OCTOBER
1957
VOLUME 21
NUMBER 5

JOURNAL OF CONSULTING PSYCHOLOGY

AMERICAN PSYCHOLOGICAL ASSOCIATION

JOURNAL OF CONSULTING PSYCHOLOGY

Edited by Laurance F. Shaffer
Teachers College, Columbia University

Advisory Editors: Donald W. Fiske, *University of Chicago.* Ann Magaret Garner, *Lincoln, Nebraska.* Harrison G. Gough, *University of California.* William A. Hunt, *Northwestern University.* E. Lowell Kelly, *University of Michigan.* Morris Krugman, *Board of Education, New York, N. Y.* Bertha M. Luckey, *Board of Education, Cleveland, Ohio.* Fred McKinney, *University of Missouri.* Anne Roe, *New York, N. Y.* Carl R. Rogers, *University of Chicago.* Nevitt Sanford, *University of California.* Edward Joseph Shoben, Jr., *Teachers College, Columbia University.* George J. Wischner, *University of Pittsburgh.* J. R. Wittenborn, *Rutgers University.*

Arthur C. Hoffman, *Managing Editor*
Helen Orr, *Assistant Managing Editor*
Editorial Staff: Barbara Cummings, Frances H. Clark, Sadie J. Doyle, Sarah Womack

The *Journal of Consulting Psychology* is the clinical journal of the American Psychological Association. It is devoted primarily to the publication of original research relevant to psychological diagnosis, psychotherapy and counseling, personality, and the dynamics of behavior. Case studies, theoretical contributions, descriptions of clinical techniques, and discussions of the professional problems of clinical psychology also appear in its pages.

Editorial Office

Manuscripts submitted for publication, tests for review, and correspondence on editorial matters should be sent to the Editor, Laurance F. Shaffer, 525 West 120th St., New York 27, N. Y. All manuscripts should be submitted *in duplicate* to facilitate referral to the Advisory Editors.

Style of Manuscripts. Manuscripts must be prepared in the style described in the *Publication Manual of the American Psychological Association (1957 Revision)*, obtainable for $1.00 from the Washington office of the APA.

Early Publication. Authors may secure the early publication of an acceptable article in the first issue to go to press after its acceptance by arranging for the article to be printed as extra pages, increasing the number of pages that subscribers receive. Information about charges for early publication may be obtained from the Editor or from the Central Office.

Brief Reports. See the back of the *Contents* page for information on the early publication of Brief Reports.

American Documentation Institute. To save printing costs, materials supplementary to articles may be deposited with the ADI. To obtain copies, order by Document Number as cited in the reference footnote, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. The indicated fee must be remitted in advance. Make checks payable to Chief, Photoduplication Service, Library of Congress.

Reprints. Authors receive fifty reprints gratis, except for early publication articles and Brief Reports. Additional quantities may be ordered when returning proofs.

Business Office

Communications concerning subscriptions, change of address, claims for the nonreceipt of a number, advertising, and other business matters, should be sent to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D.C. Address changes must reach the subscription office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

The *Journal of Consulting Psychology* is published bimonthly in February, April, June, August, October, and December, the yearly volume comprising approximately 480 pages. The subscription per year is $8.00, foreign $8.50, single number $1.50.

Published by the American Psychological Association, Inc.
at Prince and Lemon Sts., Lancaster, Pa.
and 1333 Sixteenth St. N.W., Washington, D.C.

Entered as second-class matter, March 21, 1955, at the post office at Lancaster, Pa., under the act of March 3, 1879. Acceptance for mailing at special rate of postage provided for in Section 538, Act of February 25, 1925, authorized May 27, 1940.

© 1957 by the American Psychological Association, Inc.

JOURNAL OF CONSULTING PSYCHOLOGY

October, 1957 Vol. 21, No. 5

Contents

The Scaling of Terms Used to Describe Personality: *Arnold H. Buss and Herbert Gerjuoy* - 361

What Is Measured by the "Cannot Say" Scale of the Group MMPI?: *Arthur S. Tamkin and Isidor W. Scherer* - 370

Heterosexual Somatic Preference and Fantasy Dependency: *Alvin Scodel* - 371

Identification, Parent-Cathexis, and Self-Esteem: *Sidney M. Jourard* - 375

Perceptions of Significant Family and Environmental Relationships in Aggressive and Withdrawn Children: *David S. Shapiro* - 381

The Role-Taking Hypothesis in Delinquency: *Charles F. Reed and Carlos A. Cuadra* - 386

Deriving a Measure of Test-Taking Defensiveness: *Charles Hanley* - 391

Extraversion, Neuroticism, and Manifest Anxiety: *A. W. Bendig* - 398

F–K in a Motivated Group: *James Drasgow and W. Leslie Barnette, Jr.* - 399

The Edwards Personal Preference Schedule and Social Desirability: *Robert E. Silverman* 402

The Correlates of Manifest Anxiety in Perceptual Reactivity, Rigidity, and Self Concept: *Emory L. Cowen, Fred Heilizer, Howard S. Axelrod, and Sheldon Alexander* - 405

Barron's Ego-Strength Scale: A Replication of an Evaluation of Its Construct Validity: *Arthur S. Tamkin and C. James Klett* - 412

Configurational Analysis of MMPI Profiles of Psychiatric Groups: *Earl S. Taulbee and Boyd D. Sisson* - 413

Scoring Intelligence on the Lowenfeld Mosaic Test: *Malcolm H. Robertson* - 418

The Social Desirability Stereotype in a Hospital Population: *C. James Klett* - 419

The Rorschach Performance of Epileptic Children: *Merville C. Shaw and William M. Cruickshank* - 422

The Validity of Some Current Tests for Organicity: *James W. Parker* - 425

The Relation of the Archimedes Spiral Aftereffect and the Trail Making Test to Brain Damage in Children: *Anthony Davids, Louis Goldenberg, and Maurice W. Laufer* - 429

The Tendency of the Dörken and Kral Brain Damage Measure to Score False Positives: *Harry L. Saslow and William G. Shipman* - 434

Base Rate and the Archimedes Spiral Illusion: *Donald W. Stilson, Malcolm D. Gynther, and Boris Gertz* - 435

Psychological Test Reviews - 438

Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specifications given below.

2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.

3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.

4. Agrees not to submit the full report to another journal of general circulation.

Specifications

Brief Report. The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the author's lines, must not exceed 75 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style (1). Headings, tables, and references are avoided or, if essential, must be counted in the 75 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 75-line quota: [1]

[1] An extended report of this study may be obtained without charge from John Doe, 300 Market St., Prospect 6, Mass. (giving the author's full name and address), or for a fee from the American Documentation Institute. Order Document No. ——, remitting $—— for microfilm or $—— for photocopies.

Extended report. Because the extended report is intended for photoduplication, and is not copy to be sent to a printer, its style should differ in several ways from that of other manuscripts: (*a*) The extended report should be typed with single spacing for economy in duplication. (*b*) Tables and figures should be placed adjacent to the text which refers to them. A caption should be typed below each figure. (*c*) Footnotes should be typed at the bottom of the page on which reference is made to them. In other respects, the full report is prepared in the style specified by the *Publication Manual* (1).

Reference

1. American Psychological Association. Council of Editors. *Publication manual of the American Psychological Association (1957 rev.)*. Washington, D. C.: American Psychological Association, 1957.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

The Scaling of Terms Used to Describe Personality[1]

Arnold H. Buss
Carter Memorial Hospital, Indianapolis

and Herbert Gerjuoy
Indiana University

The terms used to describe personality have diverse origins. The majority of such words have been taken over from everyday speech and from literature. The various personality theories constitute a second source, and such terms as "castration anxiety," "rich inner life," and "approach-avoidance conflict" are associated with very different orientations to personality. Finally, our techniques of personality assessment have added another kind of jargon to the language: "coarcted," "plus-getting," etc.

With three major sources of descriptive clinical language, it is not surprising that several different terms often have the same behavioral referent. Even more confusing is the fact that one term may have several different behavioral referents. This chaotic state of affairs has led to considerable trouble when psychologists attempt to communicate their evaluations of personality to each other. The lack of communication is illustrated by a study that involved an attempt to predict flying success by analyzing test protocols. In discussing the predictions the authors wrote:

> Because at least three different psychologists evaluated each cadet, the variation in personality descriptions from one clinician to the next in the kind and degree of personality descriptions makes a thorough, systematic analysis of this kind impossible. Even when those cases on which there is unanimous agreement as to predicted outcome are considered, there is little congruence of terminology, each clinician favoring concepts related to his own frame of reference. A major problem in further analysis of data of this kind is finding an adequate system for classifying descriptive terms to make comparisons among clinicians (5, p. 489).

If such a classification system were to be developed, which words would be included? It seems plausible to select those words that have the same meaning for psychologists of various orientations and eliminate those words that do not. If psychologists had available a set of terms whose meaning had been agreed upon, a large step would have been taken in the development of a scientific personality terminology.

One method of assessing interpsychologist agreement would be to compile definitions of often-used clinical terms. This approach was adopted by Grayson and Tolman (3), who collected definitions of 50 words that appeared frequently in clinical reports. They found moderate agreement on some of the terms, but in summarizing their findings they concluded, "The most striking finding of the study is the looseness and ambiguity of the definitions of many of the terms" (3, p. 229). This research would seem to confirm the prescientific state of current personality language.

However, the Grayson-Tolman study may not be as damning of present-day terminology as it first appears. Most investigators of human behavior have found wide disparities between a subject's behavior and his report of his behavior in the laboratory. Clinical psychologists might show considerable agreement if they were given a denotative or choice type of task rather than a literary type of task. For example, if the task were to evaluate the intensity of various words along some dimension, psychologists might differ very little among themselves. A physical analogy may serve to illustrate this point. Suppose several individuals are asked to define and to rank the words "hot," "warm," "tepid," "cool,"

[1] The writers wish to thank Dr. Irma Gerjuoy, Richard Mathias, and F. L. Mills for their help in collecting and analyzing the data of this study.

and "frigid" along a dimension of heat. Their definitions would probably be different, but their rankings would probably be quite similar. Unfortunately, in the Grayson and Tolman study the subjects only defined terms and did not order them along dimensions. It seems likely that such ordering would have yielded considerably more agreement than did defining the terms. Furthermore, in most clinical situations the psychologist's denotative behavior (identifying a patient as "agitated") is more important than his literary skill (defining the term "agitated"). Therefore, a scaling procedure was chosen in an attempt to discover which personality words would be reliably scaled by clinical psychologists.

Although the choice of personality dimensions had to be arbitrary, it was felt that the personality variables chosen would be meaningful to most psychologists who study personality. Factors influencing the choice were: (*a*) previous classification systems, e.g., Murray (7), Cattell (1); (*b*) importance to clinical psychologists as indicated by their reports on patients; (*c*) minimum overlap among dimensions; and (*d*) availability of words that would cover the full spectrum of a dimension. A full discussion of the issues involved in classifying behavior and selecting response units is beyond the scope of this paper. The scaling of terms does not stand or fall on the particular dimensions of personality chosen. However, in defense of the dimensions used here, it should be noted that they differ little from those used in most classification schemata in the area of personality. Furthermore, "feedback" from clinical colleagues who were presented with these dimensions revealed ready acceptance of them.

Method

The initial step involved selecting 18 dimensions of personality along which to order descriptive terms: Verbal Hostility, Physical Hostility, Hostile Attitudes, Physical Sexuality, Sexual Attitudes, Anxiety, Mood, Guilt, Self-Esteem, Ideation, Impulsiveness, Guardedness, Flexibility, Emotional Warmth, Sociability, Autonomy, Dominance, and Ambition. Next, 18 sets of adjectives were compiled, one set for each of the personality dimensions. Terms were selected which, in the opinion of the writers, are well known to most psychologists, and an attempt was made to avoid the pitfalls of both slang and pedantry. Words were chosen that had specific rather than general referents, e.g., "solemn," "brooding," and "hopeless" were selected for the Mood dimension rather than "depressed," which is general enough to include all of the other three adjectives.

Each dimension was composed of a series of words ranging from too much of the personality attribute to too little. For example, the dimension of Self-Esteem ranged from "self-exalting" through "confident" to "self-deprecatory."

It was a simple matter to assign words to most of the dimensions, but in constituting several dimensions the tendency of some adjectives to connote more than one aspect of behavior proved to be troublesome. An example of multiple connotations in physical dimensions will illustrate the difficulty: the word "light" refers to both the dimension of illumination and the dimension of weight. Similarly, the group of words "panicky," "terrified," "agitated," and "tremulous" fall at one end of the dimension of Anxiety, whose other extreme may be described by "stolid" and "phlegmatic" on the one hand, or "overconfident," "sure," and "secure" on the other hand. The writers have found it useful to refer to this case as a Y-shaped dimension. Such Y-shaped dimensions occurred in the areas of Physical Hostility, Sexuality, Ambition, Anxiety, and Flexibility. In these instances, including both possible lists of words would have necessitated repeating part of one list. Therefore, only one list was chosen for each of these five dimensions, the list that seemed to cover a wider range of behaviors of interest to clinical psychologists.[2]

Two scalings of these adjectives were performed.

[2] The writers' opinion was not the sole determiner of inclusion of words or composition of the dimensions. The results of a pilot study were also considered. In this study, 12 psychologists rated words on 16 of the 18 final dimensions. Modifications were made where there was poor interjudge agreement, and suggestions by the judges were taken into account when compiling the final list of adjectives.

Intensity Scaling

There were 42 judges[3] who were clinical psychologists with at least an MA or equivalent and one year of experience. Half the judges were recruited locally, and the other half were distributed throughout the country. The group manifested a good deal of diversity in training, theoretical orientation, interests, and age. This heterogeneity suggests that systematic bias in rank ordering the adjectives was minimal.

The judges were given the following instructions:

> Enclosed are 18 packets of adjectives, each packet being a dimension of personality. The top card names the dimension, and the adjectives are below it. The idea is to rank the adjectives according to where you think they belong on the continuum. The continua are set up so that abnormality is at *both* extremes, with normality somewhere in the middle. This differs from the usual continuum of good to poor adjustment. For example, on the Autonomy continuum the word Assured should be somewhere in the middle, Defiant somewhere near the top (too much Autonomy), and Clinging somewhere near the bottom (too little Autonomy). Thus, each dimension ranges from too much of the variable through a moderate degree to too little of it. Many of the words have connotations that have relevance for several dimensions; try to rank them with respect to only the other words on the dimension.

The rankings were converted into scale values by means of Hull's method (4), which assumes a normal distribution underlying the obtained ranks. Each ranking was converted into a percent position, which was translated into a scale value along the base line of a normal distribution curve. Then the scale values of all the judges were averaged to yield a mean scale value for each word. The standard deviation of the normalized rankings of each word was also computed.

The Intensity scaling also involved abnormality, since the judges were instructed that abnormality corresponded to too much or too little of the particular variable. However, it was not known whether the judges followed the instructions, and the relationship between intensity and abnormality remained unknown. Therefore, the words were scaled again, this time for abnormality.

Abnormality Scaling

There were 24 judges, all recruited locally, and the same minimum qualifications for judges were used as in the first scaling. Twelve of these judges had also participated in the first (Intensity) scaling, but seven months elapsed between scalings. Again there was considerable heterogeneity in training and orientation of judges.

A rating scale method was used, and judges were given the following instructions:

> The purpose of this study is to scale words on the dimension of abnormality. This dimension may be thought of in terms of a continuum of numbers that ranges from Normality at 1 to Abnormality at 9. Your task is to judge each word with respect to this dimension of abnormality. Simply assign a number between 1 (most normal) and 9 (most abnormal) to each word. The task does *not* involve ranking the words but only assigning each one to a point on the 1 to 9 abnormality continuum. There are 18 sets of words, each set covering an area of personality. Within each set the words are arranged in alphabetical order. Use each number (1 to 9) as often as you need.

Average scale values were obtained for each adjective directly by computing the arithmetic mean of the numbers assigned to it by the various judges.[4] Both the method of scaling and the kind of dimension (abnormality at the opposite extreme from normality) were different than they were for the Intensity scaling. It may be argued that the two different instructions and scaling procedures constitute a check on the stability of the scale values, since in scaling for Intensity both the instructions and the scaling method are sources of variance. Scale values should transcend the particular instructions and scaling

[3] The writers wish to thank these 42 psychologists, as well as those who served as judges in the pilot study and those in the second scaling operation. This study could not have been carried out without their considerable efforts.

[4] The raw data were amenable to scaling by the method of successive intervals (2). This method assumes that the distributions of judgments are normal on the psychological continuum. The method of scaling used here does not make this assumption, although it does assume that the nine categories of rating represent equal intervals in the psychological continuum. With well-trained raters (psychologists) it was felt that the latter assumption was reasonable, and it was desired to perform the two scalings with procedures that differed as much as possible, to provide the most stringent test of the stability of the scale values.

method used; otherwise, their applicability would be extremely limited. A high relationship between the Intensity scale values and the Abnormality scale values, despite the procedural differences involved in scaling, would suggest that the scale values have some generality.

Scale Values

The results of the two scalings are presented in Table 1. Low Intensity means correspond to an insufficiency of the personality variable, and high Intensity means correspond to an excess. The Abnormality means are directly related to abnormality: the larger the mean, the greater the degree of abnormality. Within each personality dimension the words are listed in order of their Intensity means, from lowest to highest. Inspection of Table 1 reveals that as the Intensity means increase in magnitude, the Abnormality means proceed from high values to low values to high values again. This relationship is best illustrated by a plot of the Abnormality means vs. the Intensity means, and Fig. 1 shows such a plot for the dimension of Impulsiveness. The two segments of the curve form a rough V. The left (descending) segment represents words with low to moderate Intensity values; here the Abnormality values descend from the extreme toward minimal Abnormality. The right (ascending) segment represents adjectives with moderate to high Intensity values; here the Abnormality means ascend from minimal Abnormality toward the extreme. The apex of the V corresponds to the word with the minimal Abnormality score.

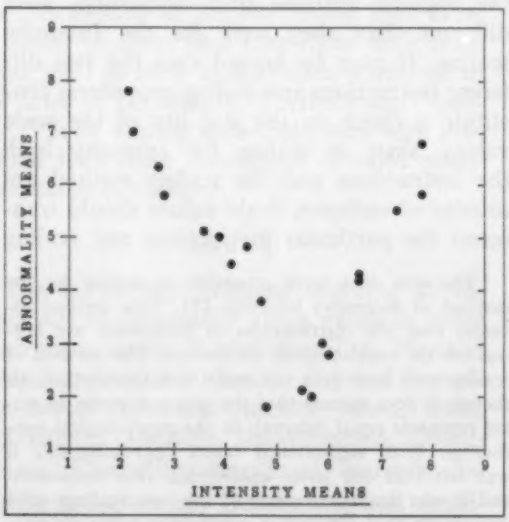

Fig. 1. A plot of abnormality means versus intensity means for the impulsiveness dimension.

In determining the strength of the relationship between the Intensity and Abnormality means, it was decided to break the curves at the apex of the V into two segments, one for the descending and one for the ascending arm of the V. Within each segment a rank-order correlation between the Intensity and Abnormality means was computed.[5] Thus, two correlations were obtained for each personality dimension, one for each portion of the V-shaped curve. These two correlations were averaged, weighting each correlation by the number of words in its portion of the V-shaped curve. This average was taken as the measure of the over-all strength of the relationship between the Abnormality and the Intensity means in each dimension.

These average rank-order correlations ranged from .75 to .97, indicating a strong relationship between Intensity and Abnormality scale values for all the dimensions. Evidently, the judges did consider abnormality when scaling for Intensity. Thus, the scale values appear to have considerable generality.

Reliability

Reliability coefficients of the mean scale values, r_{kk}, were computed for each dimension. All the Intensity reliability coefficients were above .99, and the Abnormality reliabilities ranged from .93 to .98.

The *SD*'s in Table 1 are related to these reliability coefficients in the following way: for the Intensity scaling the higher the reliability, the lower the average variance. The extremely high reliability coefficients mean that the obtained *SD*'s are extremely low.

The Abnormality *SD*'s shown in Table 1 are slightly larger than the Intensity *SD*'s. There are three reasons for this discrepancy.

[5] In order to avoid inflating the obtained correlation coefficients, the pair of mean scale values for that adjective with minimal Abnormality score on any given dimension was omitted from the computations.

Table 1

Mean Scale Values and *SD*'s for the Intensity and Abnormality Scalings

	Intensity		Abnormality	
	Mean	*SD*	Mean	*SD*
Verbal Hostility				
1. venomous	1.7	.63	7.7	2.08
2. abusive	2.3	.54	6.9	1.50
3. threatening	2.6	.92	6.6	1.95
4. derisive	3.4	.53	5.8	1.94
5. derogatory	3.5	.51	5.5	1.98
6. scornful	3.7	.59	5.5	1.71
7. sarcastic	4.1	.43	5.4	1.61
8. argumentative	4.4	.53	4.8	1.31
9. over-critical	4.5	.35	5.0	1.70
10. nagging	4.6	.31	5.5	1.71
11. outspoken	5.1	.24	3.1	1.44
12. frank	5.5	.20	2.5	1.38
13. tactful	5.7	.15	1.8	1.14
14. soft-spoken	6.0	.38	2.8	1.43
15. complimentary	6.3	.24	3.1	1.61
16. praising	6.7	.37	3.8	1.76
17. flattering	7.0	.43	5.0	1.44
18. mealy-mouthed	7.2	1.51	5.9	2.01
19. apple-polishing	7.6	.66	6.0	1.51
20. eulogistic	8.1	.88	5.5	1.85
Physical Hostility				
1. murderous	1.7	.18	8.8	.64
2. assaultive	3.2	.37	8.0	1.34
3. destructive	3.4	.52	7.0	1.67
4. combative	3.9	.36	6.6	1.63
5. hot-blooded	4.4	.18	5.3	1.67
6. even-tempered	5.1	.26	1.5	.91
7. peaceable	5.7	.34	2.1	1.09
8. harmless	6.1	.71	4.1	1.79
9. inhibited	6.4	.72	4.7	1.60
10. placating	6.9	.53	4.2	1.60
11. cringing	8.2	.41	7.1	1.61
Hostile Attitudes				
1. malicious	1.6	.60	6.9	1.72
2. embittered	2.8	.86	5.8	1.74
3. quarrelsome	2.9	.80	5.8	1.63
4. surly	3.0	.80	6.6	1.80
5. provocative	3.5	.96	5.9	1.68
6. resentful	3.8	.59	5.6	1.82
7. irritable	4.3	.56	4.8	1.51
8. grouchy	4.3	.54	4.8	1.55
9. petulant	4.3	.74	5.6	1.52
10. grudging	4.3	.62	5.2	1.58
11. civil	5.4	.38	2.4	1.19
12. inoffensive	5.7	.53	3.6	1.44
13. unresentful	5.8	.62	3.6	1.73
14. agreeable	6.0	.68	1.6	.90
15. gentle	6.4	.64	2.3	1.18
16. gracious	6.5	.76	2.2	1.20
17. conciliatory	6.6	.65	3.0	1.31
18. ingratiating	7.3	.81	5.0	1.83
19. oily	7.7	1.05	6.5	1.63
20. fawning	7.8	1.09	6.1	1.63

	Intensity		Abnormality	
	Mean	*SD*	Mean	*SD*
Physical Sexuality				
1. assaultive	1.8	.53	8.4	1.11
2. wanton	2.8	.71	7.6	1.68
3. over-active	3.6	.49	5.7	1.65
4. soliciting	3.8	.85	6.2	1.35
5. passionate	3.9	.57	2.8	1.68
6. assertive	4.3	.56	3.2	1.59
7. amorous	4.6	.42	2.6	1.70
8. permissive	5.1	.44	2.9	1.86
9. hesitant	5.8	.36	4.3	1.64
10. passive	6.0	.60	4.8	1.84
11. restrained	6.4	.63	4.1	1.76
12. inhibited	6.8	.41	5.6	1.49
13. apathetic	7.3	.87	6.8	1.55
14. abstinent	8.1	.62	6.6	2.00
Sexual Attitudes				
1. lewd	2.5	1.18	7.1	2.04
2. lustful	2.5	.70	6.0	2.01
3. erotic	3.6	.54	3.6	2.14
4. sensual	3.8	.43	3.2	1.81
5. romantic	4.5	.17	2.2	1.14
6. unsure	5.0	.21	4.1	1.81
7. prim	5.7	.34	5.4	1.80
8. prudish	6.3	.66	6.0	1.84
9. ashamed	6.6	.81	5.9	1.73
10. puritanical	7.0	.87	7.1	1.78
11. disgusted	7.8	.74	7.3	1.46
Anxiety				
1. terrified	1.8	.54	8.4	1.63
2. panicky	2.2	.54	7.8	1.29
3. agitated	2.9	.53	7.0	1.17
4. tremulous	3.5	.41	6.9	1.74
5. apprehensive	3.8	.43	5.2	1.56
6. tense	4.1	.30	5.3	1.55
7. fretful	4.3	.30	5.1	1.55
8. uneasy	4.6	.26	4.2	1.75
9. composed	5.3	.44	1.5	.71
10. calm	5.5	.53	1.4	.70
11. nonchalant	6.0	.61	3.1	1.52
12. unconcerned	6.1	.45	4.1	1.82
13. cool	6.1	.94	2.3	1.34
14. bland	6.7	.93	4.8	1.55
15. stolid	7.2	.89	5.2	1.61
16. imperturbable	7.3	.88	3.8	1.58
17. phlegmatic	7.7	.90	5.4	1.89
Mood				
1. euphoric	1.3	.38	6.7	1.82
2. elated	2.2	.40	4.1	1.94
3. frivolous	2.8	.73	5.1	1.55
4. buoyant	3.2	.60	3.1	1.54
5. gay	3.4	.42	2.7	1.21
6. jovial	3.5	.41	2.1	1.13
7. light-hearted	4.0	.31	2.4	1.32

Table 1—*Continued*

	Intensity		Abnormality	
	Mean	*SD*	Mean	*SD*
8. cheerful	4.1	.27	1.4	.63
9. placid	4.5	.10	3.3	1.95
10. sober	4.9	.20	3.4	1.65
11. serious	5.0	.25	3.2	1.61
12. solemn	5.4	.24	4.5	1.89
13. mirthless	5.5	.64	5.8	1.71
14. grave	5.8	.35	5.3	1.93
15. gloomy	6.6	.54	5.6	1.63
16. brooding	6.7	.70	5.8	1.64
17. dejected	7.0	.52	5.7	1.54
18. disconsolate	7.1	.68	6.9	1.55
19. despondent	7.6	.65	7.0	1.51
20. hopeless	8.4	.68	7.9	1.26
Guilt				
1. self-condemning	1.6	.49	7.4	1.55
2. self-reproachful	2.7	.47	5.7	1.77
3. remorseful	2.8	.75	4.2	2.18
4. ashamed	3.3	.44	3.4	1.82
5. chagrined	3.7	.56	3.7	1.62
6. regretful	3.9	.27	3.5	2.00
7. concerned	4.4	.10	2.0	1.43
8. indifferent	4.8	.52	5.9	1.71
9. unfeeling	5.6	.91	6.7	1.84
10. unreformed	5.7	.49	5.8	1.73
11. cynical	5.8	.78	5.5	1.78
12. unrepentant	5.9	.58	5.8	1.88
13. hardened	5.9	.82	6.0	1.65
14. shameless	6.5	.69	6.7	1.46
15. conscienceless	7.0	.95	8.2	.88
16. unscrupulous	7.7	.85	7.2	1.69
17. incorrigible	7.7	.85	8.0	1.58
Self-Esteem				
1. self-exalting	2.0	.88	6.7	1.84
2. pompous	2.6	.86	5.8	1.95
3. conceited	3.4	.57	5.4	1.77
4. boastful	3.4	.63	4.9	1.71
5. vain	3.5	.58	5.8	1.86
6. cocky	3.6	.85	4.7	1.86
7. confident	4.6	.12	1.3	.69
8. self-respecting	4.7	.13	1.3	.69
9. modest	5.3	.22	2.4	1.44
10. unassuming	5.3	.23	2.4	1.22
11. humble	5.9	.32	4.0	1.79
12. self-doubting	6.4	.45	5.9	1.55
13. self-effacing	6.8	.66	6.2	1.71
14. self-deprecatory	7.4	.71	7.3	1.74
15. forlorn	7.5	.96	6.9	1.74
16. self-abasing	7.7	.73	7.5	1.80
Ideation				
1. delusional	1.8	.73	8.8	.47
2. ruminative	3.1	.94	5.3	1.59
3. day-dreaming	3.4	.52	4.8	1.71
4. fanciful	3.6	.57	4.5	1.91
5. musing	4.3	.44	3.0	1.46
6. contemplative	4.4	.64	2.1	1.19
7. thoughtful	4.9	.35	1.6	1.04
8. matter-of-fact	5.6	.33	3.4	1.58
9. literal	5.9	.64	4.2	1.73
10. unreflective	6.4	.56	5.5	1.61
11. unimaginative	6.9	.58	5.5	1.66
12. stolid	6.9	.72	5.0	1.81
13. vacuous	8.1	1.26	7.5	1.44
Impulsiveness				
1. incontinent	2.1	1.04	7.8	1.31
2. reckless	2.2	.83	7.0	1.78
3. rash	2.7	.62	6.3	1.79
4. impetuous	2.8	.58	5.8	1.98
5. excitable	3.6	.57	5.1	1.99
6. hasty	3.9	.36	5.0	1.70
7. abrupt	4.1	.51	4.5	1.55
8. restless	4.4	.40	4.8	1.67
9. mobile	4.7	.28	3.8	1.78
10. spontaneous	4.8	.40	1.8	1.14
11. self-possessed	5.5	.64	2.1	1.67
12. cool-headed	5.7	.53	2.0	1.02
13. deliberate	5.9	.42	3.0	1.47
14. controlled	6.0	.61	2.8	1.45
15. restrained	6.6	.36	4.2	1.53
16. staid	6.6	.54	4.3	1.51
17. over-cautious	7.3	.64	5.5	1.85
18. retarded	7.8	.87	6.8	1.39
19. sluggish	7.8	.79	5.8	1.91
Guardedness				
1. uncommunicative	1.9	.57	6.5	2.18
2. secretive	2.1	.71	6.5	1.35
3. taciturn	3.4	.62	4.9	2.04
4. close	3.6	.61	4.4	1.50
5. reticent	3.8	.58	4.6	1.47
6. noncommittal	4.0	.66	4.5	1.73
7. reserved	4.3	.48	3.7	1.57
8. discreet	4.6	.45	2.0	1.06
9. straightforward	5.6	.62	2.3	1.46
10. candid	5.9	.77	2.7	1.91
11. open	6.0	.86	2.9	1.76
12. unreserved	6.2	.82	4.4	2.10
13. talkative	6.4	.46	3.8	1.67
14. voluble	7.3	.93	5.7	2.13
15. verbose	7.3	.77	5.7	1.81
16. garrulous	7.6	1.00	5.4	1.95
Flexibility				
1. spineless	1.8	1.11	7.3	1.75
2. over-suggestible	2.6	.53	6.3	1.77
3. yielding	3.5	.54	4.3	1.62
4. docile	3.6	.73	5.2	1.52
5. changeable	3.7	.53	3.5	1.44
6. amenable	4.4	.32	2.0	1.02

Table 1—*Continued*

	Intensity		Abnormality	
	Mean	*SD*	Mean	*SD*
7. adaptable	4.5	.44	1.2	.47
8. conventional	4.9	.31	4.0	1.37
9. persistent	5.5	.37	4.0	1.71
10. habit-bound	5.9	.60	6.0	1.31
11. stubborn	6.3	.57	5.1	1.86
12. perseverating	6.7	1.11	7.0	1.54
13. unbending	6.7	.63	6.0	1.59
14. mulish	7.1	1.01	7.1	1.44
15. intractible	7.9	.78	7.2	1.41
Emotional Warmth				
1. over-indulgent	2.0	.73	5.8	1.58
2. doting	2.2	.57	5.4	1.65
3. affectionate	3.5	.62	1.2	.52
4. sentimental	3.5	.57	3.8	1.46
5. tender	3.8	.68	2.1	1.00
6. sympathetic	4.2	.58	2.4	1.15
7. kindly	4.6	.36	1.7	.90
8. considerate	4.8	.31	1.5	.76
9. cool	5.8	.51	4.6	1.87
10. unresponsive	5.9	.60	6.7	1.77
11. detached	6.0	.73	5.8	1.69
12. unfeeling	6.5	.57	7.1	1.48
13. hardened	6.7	.62	6.9	1.66
14. rejecting	7.7	.95	7.2	1.29
15. frigid	7.8	.79	7.9	1.12
Sociability				
1. intrusive	2.0	.86	5.5	1.61
2. meddlesome	2.1	.97	6.3	1.82
3. gregarious	2.9	.67	2.1	1.17
4. convivial	3.3	.58	1.9	.95
5. intimate	3.4	.75	2.9	1.41
6. comradely	3.8	.37	1.7	.85
7. companionable	4.0	.38	1.5	.64
8. agreeable	4.4	.25	1.7	.84
9. accessible	4.8	.72	2.3	1.38
10. hesitant	5.2	.29	4.2	1.47
11. reserved	5.3	.47	4.5	1.66
12. bashful	5.6	.32	4.3	1.70
13. reticent	5.6	.39	5.2	1.75
14. retiring	5.9	.27	5.4	1.75
15. shrinking	6.5	.53	7.0	1.57
16. seclusive	7.2	.65	7.1	1.51
17. withdrawn	7.5	.75	8.3	1.25
18. solitary	7.5	.80	7.5	1.66
19. isolated	7.8	.73	7.0	1.94
Autonomy				
1. rebellious	1.8	.71	6.5	1.50
2. defiant	2.1	.63	6.3	1.84
3. negativistic	3.0	.69	6.5	1.76
4. contrary	3.6	.53	5.5	1.78
5. resistant	3.7	.60	4.5	1.47
6. wilful	3.7	.63	4.7	1.75
7. dissenting	4.1	.60	4.6	1.71
8. self-reliant	4.6	.51	1.5	1.00
9. assured	4.6	.57	1.5	1.08
10. resigned	5.5	.67	5.4	2.12
11. weak	5.9	.73	6.1	1.44
12. quitting	6.0	1.10	6.1	1.83
13. plaintive	6.1	.53	5.6	1.68
14. suppliant	6.1	.78	5.3	1.62
15. pleading	6.7	.93	5.9	1.86
16. whining	6.9	.67	7.0	1.81
17. clinging	7.4	1.00	6.4	1.80
18. helpless	7.8	1.14	7.8	1.71
Dominance				
1. dictatorial	1.6	.41	7.7	1.52
2. autocratic	2.4	.43	6.7	1.27
3. high-handed	3.1	.56	6.8	1.55
4. masterful	3.4	.50	2.9	1.11
5. forceful	3.7	.37	4.4	1.78
6. assertive	4.0	.42	2.5	1.38
7. decisive	4.2	.24	2.4	1.32
8. cooperative	4.8	.32	1.8	1.09
9. assenting	5.3	.42	4.1	1.53
10. conforming	5.5	.52	4.2	1.77
11. compliant	5.9	.55	4.4	1.49
12. acquiescent	5.9	.65	4.7	1.77
13. imitative	6.2	.82	5.4	1.53
14. deferent	6.4	.76	4.8	1.61
15. timid	6.9	.47	5.6	1.84
16. meek	7.3	.59	5.6	1.87
17. servile	8.5	.34	7.5	1.53
Ambition				
1. grandiose	1.5	.00	7.5	1.83
2. pretentious	2.7	.58	5.9	1.72
3. aspiring	3.5	.51	1.8	.76
4. enterprising	3.7	.47	1.8	1.26
5. persistent	3.8	.75	2.9	1.44
6. eager	3.8	.50	2.2	1.18
7. content	4.9	.24	2.5	1.58
8. self-satisfied	4.9	.25	3.5	1.71
9. complacent	5.3	.33	3.9	1.66
10. lackadaisical	6.0	.40	5.9	1.81
11. indifferent	6.1	.67	5.5	1.91
12. listless	6.6	.74	6.0	1.92
13. indolent	7.0	.96	6.4	1.87
14. apathetic	7.6	1.00	6.8	1.98
15. lethargic	7.6	.82	6.7	1.71

First, the Abnormality reliabilities were lower. Second, the number of judges on which the scale values were based was smaller. Third, the Abnormality *SD*'s were inflated by differences among judges in average rating, whereas the Intensity *SD*'s could not be affected by

this variable because the rank-order method was used.

Discussion

The results clearly indicate that clinical psychologists can order descriptive terms consistently. In both the Intensity and Abnormality scalings the vast majority of personality terms were ordered with relatively low interjudge variability. It may be concluded that when the task is denotative and the personality dimensions are specified, many words in current usage have behavioral referents clear enough for judges to agree in scaling them.

These findings may be extrapolated to the written reports of clinical psychologists. The clinician's predilection for using terms dear to his theoretical or test orientation must be assessed in the light of the communication value of such terms. Words that are scaled less consistently should be omitted, and words with lower interjudge variability should be included in psychological evaluations. In addition, the dimension of personality should be specified when descriptive terms are used. For example, the word "cool" may refer to a relative lack of anxiety or to a relative lack of emotional warmth; but when the dimension is specified, the correct connotations are communicated. When the personality dimension is not stated, communication often breaks down.

In addition to suggesting which terms to use in evaluations of personality, the present classification system may be a useful tool for personality research. It might be utilized in an assessment study of the Holtzman and Sells type (5). Words could be selected on the basis of Intensity or Abnormality. Clinicians usually prefer to use words rather than numbers in characterizing personality. With the present classification system they could choose words to describe personality, whose scale values for Intensity and Abnormality had already been determined. The question of reliability among clinicians in assessing personality (a thorny issue in the Holtzman and Sells study) could be settled easily, since all clinicians would choose words from the same lists and the differences among clinicians could be determined quantitatively.

An investigator planning to use the Intensity scale values would probably not need the entire list of words on any dimension. The following suggestion may prove helpful in selecting words. If two words have similar or identical scale values, select the word with the smaller Intensity standard deviation. If there is still doubt, use the Abnormality standard deviations to decide which word to include. Similarly, if the investigator is primarily interested in the Abnormality scale values, he would use the Abnormality standard deviations first in breaking ties in Abnormality mean scale values.

The classification system can be used as a self-rating device. Zuckerman *et al.* (8) used the data of the pilot study to select terms along 16 dimensions of personality. They had patients and normals describe themselves, their ideal, their parents, and others, and the various discrepancies between self and ideal, self and parents, etc., were investigated. What distinguished their study from most similar studies was that they were able to specify in which areas of personality there were large discrepancies.

Mills [6] used the present data to investigate personality factors in judging the behavior of others. Both the judges and the individuals who were judged assessed themselves by checking one word on each of the 18 personality dimensions. Both Mills and Zuckerman *et al.* selected words on the basis of Intensity scale values, spacing terms so as to cover the whole dimension in more or less equal intervals. The more difficult terms had to be eliminated so that the vocabulary level did not exceed that of the noncollege population. In both investigations, the time taken by subjects to complete the assessment of self and of others was considerably less than that required by a *Q* sort.

The present study has implications that go beyond the possible applications of the classification system. The language of clinical psychology is largely prescientific and imprecise. In the search for a precise and objective terminology, it would seem that psychological scaling methods have much to offer.

Summary

In this study, adjectives were scaled along 18 dimensions of personality, using rank-or-

[6] Unpublished study.

der and rating-scale methods. The terms were scaled for Intensity (too much to too little of the variable) and for Abnormality (minimal to maximal abnormality); there were 42 and 24 judges, respectively, all clinical psychologists. The mean scale values proved to be reliable, and possible applications of them were mentioned. It was suggested that psychological scaling methods are of service in improving personality terminology.

Received December 21, 1956.

References

1. Cattell, R. B. The description of personality: basic traits resolved into clusters. *J. abnorm. soc. Psychol.*, 1943, **38**, 476–506.
2. Edwards, A. L. The scaling of stimuli by the method of successive intervals. *J. appl. Psychol.*, 1952, **36**, 118–122.
3. Grayson, H. M., & Tolman, Ruth S. A semantic study of concepts of clinical psychologists and psychiatrists. *J. abnorm. soc. Psychol.*, 1950, **45**, 216–231.
4. Guilford, J. P. *Psychometric methods*. New York: McGraw-Hill, 1936. Pp. 248–249.
5. Holtzman, W. H., & Sells, S. B. Prediction of flying success by clinical analysis of test protocols. *J. abnorm. soc. Psychol.*, 1954, **49**, 485–490.
6. Horst, P. A generalized expression for the reliability of measures. *Psychometrika*, **14**, 21–31.
7. Murray, H. A., *et al. Explorations in personality*. New York: Oxford Univer. Press, 1938.
8. Zuckerman, M., Baer, M., & Monashkin, I. Acceptance of self, parents and people in patients and normals. *J. clin. Psychol.*, 1956, **12**, 327–332.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

What Is Measured by the "Cannot Say" Scale of the Group MMPI?[1]

Arthur S. Tamkin and Isidor W. Scherer
Veterans Administration Hospital, Northampton, Massachusetts

While a large fund of information is available about the various MMPI clinical and validity scales and the personality attributes which they purport to measure, relatively little has been written about the "Cannot Say" scale. The purpose of this study is to determine if the "Cannot Say" scale of the group form is the reflection of a defensive or evasive attitude on the part of the examinee, as suggested by Brown (1), and if it is related to the symptoms of psychasthenia and depression in psychiatric patients, as observed by the test authors (2).

Completed successively administered MMPI answer sheets were obtained from 126 male psychiatric patients who were newly admitted or readmitted for hospital treatment. They were scored for the "Cannot Say" scale, the three validity scales, and the nine clinical scales. The frequency distribution of the "Cannot Say" scores was described by a discontinuous curve, the major portion of which showed marked positive skewness. The scores ranged from zero to 145, the mode being zero and the median one. The discontinuous portion of the curve was made up of 12 extreme cases whose scores ranged from 26 to 145.

In order to explore the hypothesis that "Cannot Say" scores may be related to defensive attitudes or to the symptoms of psychasthenia and depression, the relationship between "Cannot Say" and each of the following scales was evaluated by chi-square tests: *K*, *L*, *F*, *Pt*, and *D*. None of the chi squares reached significance at the .05 level or better, however. The Kolmogorov-Smirnov test failed to show that subjects omitted items primarily on the psychiatric scales as compared to the remainder of the MMPI. When the hypothesis that the 12 extreme cases omitted a disproportionately larger number of items on particular psychiatric scales was tested, only one scale, *Ma*, reached the .05 level of confidence. It was concluded that for the group form of the MMPI there is no relationship between high "Cannot Say" scores and MMPI measures of depression and psychasthenia, and that high "Cannot Say" scores, as a general rule, do not seem to represent a defensive, evasive attitude on the part of the psychiatric patient. This, however, does not rule out the possibility that certain individual patients may utilize omissions as an evasive measure, but merely makes untenable the general application of such an hypothesis.

Brief Report.
Received March 15, 1957.

[1] An extended report of this study may be obtained without charge from Arthur S. Tamkin, Veterans Administration Hospital, Northampton, Mass., or for a fee from the American Documentation Institute. Order Document No. 5273, remitting $1.25 for microfilm or $1.25 for photocopies.

References

1. Brown, N. M. Evaluating and scoring the Minnesota Multiphasic "Cannot Say" items. *J. clin. Psychol.*, 1950, 6, 180–184.
2. Hathaway, S. R., & McKinley, J. C. *Minnesota Multiphasic Personality Inventory.* Manual. (Rev. ed.) New York: Psychological Corp., 1951.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

Heterosexual Somatic Preference and Fantasy Dependency

Alvin Scodel
The Ohio State University

With the exception of occasional statements in the psychoanalytic journals, almost no attention has been given to the female body-types considered desirable by adult males, and the psychological correlates of such preferences. It is, indeed, rather curious in light of the numerous attempts to view all behavior and attitudes from Freudian assumptions that a topic which so preempts the conversation and fantasies of adult males should have been so completely ignored for systematic psychological investigation.

When somatic preferences are discussed in the psychological literature, either in case studies or as part of an exposition of a particular theoretical position, inconsistent conclusions are often reached. Tridon (8) asserts that men nursed at the breast in infancy are attracted in adulthood to women with well developed breasts, and that men who have been fed by bottle prefer thin, boyish looking girls. Breast feeding, presumably a more gratifying experience than bottle feeding, leads to a later preference for women with well developed breasts who can continue to supply the oral gratifications of infancy. On the other hand, Gorer (4) states that the fetish that exists for breasts in American culture is a derivative of scheduled feeding in infancy. In this view, then, a preference for large breasted females is a result of oral frustration rather than oral satisfaction.

Despite such differences, there is little disagreement about the importance of the breast to the concept of orality. In Freud's opinion (2), the child at the breast formed the prototype of all future love relationships. Illustrative clinical illustrations can be found in cases described by Gero (3), and Levey (6). Gero's patient had been separated from his mistress, and it is stated that "his longing for this woman was not a man's longing for a lost love, but a child's longing for his mother. . . . Whenever he thought of his beloved, the aim of his yearning, the lost paradise in his imagination was always the breast" (3, p. 448). Levey discussed an ulcer patient who "fantasies the earth, the air, and the universe as composed of breasts, himself floating in a sea of breasts, and the Capitol at Washington as having a breast pinned on it" (6, p. 163). It seems to be clearly implied that the orally fixated person is preoccupied with the ample, well developed breast rather than the small breast. By way of a casually stated *ipse dixit* common clinical parlance also speaks of large breast preference as a product of oral fixation.

The present investigation is concerned with the relationship of breast size preference to an allegedly important aspect of orality; namely, dependency. The phrases "oral passivity" and "oral dependence" are so frequent in both the psychological and psychoanalytic literature that no documentation is necessary. The instrument used to measure dependency in the present study is the TAT. To be more precise, fantasy rather than overt dependency is being measured, although there is evidence to indicate that fantasy dependency as measured by the TAT correlates significantly with overt dependency when overt dependency is measured by conformity to erroneous group judgments (5).

Procedure

Five male graduate students rated 101 full-length pictures of nude females (obtained from photography and art magazines) along a dimension of breast size. The following rating scale was used:

1. Considerably above average breast size.
2. Slightly above average breast size.
3. Average breast size.
4. Slightly below average breast size.
5. Considerably below average breast size.

By summing the judges' ratings, a range of judgments for all pictures was obtained which extended from five to 25. Those pictures with a total rating of 11 or less were defined as large breasted (LB), those with a total rating of 19 or more as small breasted (SB). Sixty-five of the pictures met these criteria. Of these 65, 32 were classified as LB, and 33 as SB. Sixty-eight per cent agreement was achieved by the judges in their ratings of the 65 pictures. Ninety per cent agreement was obtained by combining Categories 1 and 2 and Categories 4 and 5.

In order to obtain a higher probability that an experimental subject would select between two females on the basis of breast size, it was necessary to equate the pictures as much as possible in general attractiveness. Otherwise, a subject who usually preferred LB females might on occasion select an SB female because of her greater general appeal. Similarly, the pairing of an attractive LB female with an unattractice SB female could lead a subject to select the former although he generally preferred SB females. To minimize the effect of facial attractiveness, masking tape was placed over the face on each of the pictures. Five additional judges then rated each picture on a six-point scale ranging from 1 (exceptionally attractive) to 6 (exceptionally unattractive). The range of judgments was eight to 23. The purpose of obtaining these ratings at this time was to determine whether a sufficiently large number of pictures classified as LB and SB would fall within a similar range of attractiveness in order to permit pairing. The distribution of ratings on attractiveness is presented in Table 1.

Table 1

Distribution of Judges' Ratings on Attractiveness

Sum of Five Judges' Ratings	Number LB	Number SB
8–12	13	0
13–20	17	19
21–23	2	14
	32	33

It is apparent that LB females were rated as more attractive. The next step consisted of having slides made out of the pictures. Forty-nine of the pictures were converted into black and white 2×2 slides with the facial area still masked out. Since conversion of the pictures into slides could conceivably have some effect upon general attractiveness, inasmuch as color, shading, and lightness and darkness were changed to a degree, a third group of judges, four in number, rated the projected slides on the same six-point scale of attractiveness used previously. A range of ratings was obtained from eight to 19. For the final sample, 20 slides were selected, 10 LB and 10 SB, all falling within the range of 12 to 15 on attractiveness. The average of the 10 LB slides was 13.2, for the SB, 13.3. An effort was made to match the slides as much as possible with respect to the position of the woman, i.e., to have both members of a pair standing or kneeling, both in profile or front view, etc.

In order to help disguise the purpose of this part of the study, 10 additional pairs, or a total of 20 pairs, were presented to the subjects. The 10 pairs discussed in the previous paragraph were the only ones used to select subjects on the variable of somatic preference. Of the 10 additional pairs, five were of two LB females, and the other five of two SB females. A table of random numbers was used to determine the order of presentation of the 20 pairs. A crucial pair was presented on the 1st, 6th, 8th, 9th, 13th, 14th, 15th, 17th, 18th, and 19th trials. For each crucial pair, the order of presentation of the LB and SB slides was alternated throughout the series.

The subjects were 169 male students enrolled in an introductory psychology course at Ohio State University. They were seen in small groups ranging in number from five to 15.

At the beginning of the experimental session, seven TAT cards intended to elicit dependency themes were administered. The cards chosen for this purpose were 4, 6BM, 7BM, 12, 13B, 14, and 18BM. The test was presented as one of creative imagination which

the examiner was attempting to standardize (the TAT had not been discussed in the course yet), and subjects were given five minutes to write each story. Aside from these departures, the usual TAT instructions were given. At the conclusion of the TAT administration, and following the distribution of answer sheets, these instructions were read:

> As you may know from your psychology course, Professor Sheldon of Harvard has done considerable research on the different kinds of body-types. We are interested in following up some of his research by finding out what female body-types are generally preferred by the typical American male. One reason for this research is that the kinds of body-types preferred by American men may have some effect on their choice of marital partners, and this in turn can affect the body-types of future generations.
>
> You will be presented with 20 pairs of slides. Each slide contains the picture of an attractive nude female. The numbers on your sheet stand for the pair number. Following each pair number is an A and a B. The A represents the first slide of the pair that is presented to you and the B represents the second slide of the pair. After looking at both slides of the pair, you are to put a circle around the A if you like the first slide better or put a circle around the B if you like the second slide better. For each pair, you encircle the letter representing the slide you prefer.
>
> You are requested not to talk or make any spontaneous exclamations during the presentation of the slides. The slides must be shown in silence.

The entire sequence of slides required about seven or eight minutes for presentation. Each slide was projected for five seconds, three seconds were allowed between the slides of a pair, and five seconds were allowed between adjacent pairs. Cooperation of the subjects was good and, with rare exceptions, little flippancy was noted. Randomly obtained comments at the end of the experimental session revealed a general lack of awareness concerning the crucial variable in this part of the study.

Results

The distribution of somatic preferences approximated a normal distribution. As previously stated, the judges showed a distinct preference for large-breasted females, but the equation of the two groups for general attractiveness insured the variability which was obtained in the preferences of subjects. Subjects who selected from 0 to 3 of the large-breasted females were defined as the SB group ($N = 28$). The LB group was comprised of those who selected from 8 to 10 of the large-breasted females ($N = 35$). Inasmuch as the mean number of LB preferences was 5.56, a middle or no-preference group was obtained by selecting those subjects with either five or six large-breasted preferences ($N = 57$).

The TAT stories of all three groups were independently scored by the writer and a graduate student in clinical psychology. Neither had any awareness of the subjects' somatic preference ratings at the time of the TAT scoring. Following the procedure of Kagan and Mussen (5), a story was scored as *D* (for dependency) if the hero sought help from another person in solving a personal problem or was disturbed over the loss of a source of love and affection. In most respects, *D* would be similar to Murray's n succorance. Only one *D* theme was scored for each story so that for any given story *D* was simply scored present or absent. Of a total of 840 stories there was agreement on 782. This 93% agreement indicates considerable reliability in scoring. A third judge arbitrated the 58 stories on which there was disagreement.

The number of *D* themes in the three groups of 120 protocols ranged from zero to six with a median of one. In both the middle and LB groups the distribution of *D* themes was markedly skewed in the positive direction. Chi square was therefore used to test for the significance of differences between groups. The results are given in Table 2. For each of the three comparisons the median value was one.

Table 2

Chi-square Differences Between SB, Middle and LB Groups in TAT *D* Themes

	TAT *D* Themes			
Groups	Above Median	At or Below Median	Chi Square	*p*
SB	18	10	6.76	.01
LB	11	24		
SB	18	10	5.70	.02
Middle	21	36		
LB	11	24	.28	.60
Middle	21	36		

The SB group gave significantly more *D* themes than either the LB or middle group. There was no significant difference between the middle and the LB groups.

Discussion

The present result is not what might have been expected on the basis of Freudian theory. In general, the theory has tended to view behaviors as the products of either excessive frustrations or excessive satisfactions (reinforcements) in an earlier period. With respect to the variable under consideration here, most analytically oriented writers who have discussed the matter at all have held that large breast preference is a result of earlier oral frustration. In Levey's discussion of the ulcer patient, for example, it is suggested that the frustrated dependency needs of the patient sought satisfaction in a constant preoccupation with large breasts. Ostensibly, such females should be perceived as potentially more nurturant for basically dependent males if the important determinant underlying such a somatic preference is frustrated dependency.

If the expression of dependency needs in TAT stories can be taken to signify conflicts in this area predicated on the frustration of such needs, and if the absence of these same needs in TAT stories implies experiential reinforcement of them, the present results are more in accord with the view that drive strength is increased by reinforcement. Large breast preference would be regarded as the consequence of continued satisfaction of dependency needs rather than their frustration. Such an interpretation is consistent with the findings of Bernstein (1). In studying the effects of infantile experiences on later activities, he raised the question of whether it was the child with the greatest amount of sucking practice as an infant who continued to use the sucking act as a preferred mode of behavior or, rather, the child with the least amount of sucking practice who continued to seek satisfaction through sucking. He found that the amount of sucking reinforcement experienced by the group whose mothers reported resistance to weaning exceeded the sucking reinforcement for the group without weaning difficulty. His results, then, support a reinforcement rather than a frustration theory of drive strength. Sears and Wise (7) are of the same view in asserting that the strength of the oral drive varies with the number of opportunities for its reinforcement so that the longer the child feeds by means of sucking, the stronger his oral drive will be. The present result is consistent with such formulations if the assumptions concerning the nature of TAT dependency can be maintained.

Summary

The purpose of this study was to ascertain possible relationships between somatic preference (preference for either large- or small-breasted females) and dependency as measured by the TAT. After writing stories to seven of the TAT cards, 169 male subjects were presented with 20 pairs of slides; 10 pairs consisted of a small-breasted and a large-breasted female, previously equated on attractiveness. On the basis of the subjects' selections, large breast preference, small breast preference, and no preference groups were elicited.

The small breast preference group gave significantly more TAT dependency themes than either of the other two groups. Speculations for this result, which is contrary to a widely held Freudian hypothesis, are offered on the basis of a reinforcement theory of learning.

Received December 6, 1956.

References

1. Bernstein, A. Some relations between techniques of feeding and training during infancy and certain behavior in childhood. *Genet. Psychol. Monogr.*, 1955, **51**, 3–44.
2. Freud, S. *Three contributions to the theory of sex.* New York & Washington: Nervous and Mental Disease Publishing Co., 1920.
3. Gero, G. The construction of depression. *Internat. J. Psychoanal.*, 1936, **17**, 423–461.
4. Gorer, G. *The American people.* New York: Norton, 1948.
5. Kagan, J., & Mussen, P. H. Dependency themes on the TAT and group conformity. *J. consult. Psychol.*, 1956, **20**, 29–32.
6. Levey, H. B. Oral trends and oral conflicts in a case of duodenal ulcer. In F. Alexander & T. M. French (Eds.), *Studies in psychosomatic medicine.* New York: Roland Press, 1948.
7. Sears, R. R., & Wise, G. W. 1. Relation of cup feeding in infancy to thumb sucking and the oral drive. *Amer. J. Orthopsychiat.*, 1950, **20**, 123–136.
8. Tridon, A. *Psychoanalysis and love.* New York: Brentano's, 1922. Reprinted in A. Tridon, *Psychoanalysis and love.* New York: Permabooks Edition, 1949.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

Identification, Parent-Cathexis, and Self-Esteem[1]

Sidney M. Jourard[2]
Emory University

Clinical observations provide the basis for suspecting that identification with parents, feelings and attitudes toward parents (parent-cathexis), and self-esteem all cohere as a syndrome. The present investigation is undertaken to verify the hypothesis of covariation among these variables, and to test relevant hypotheses concerned with identification and with the cathexis-response to persons—in this instance, parents and the self.

Method

Hypotheses

1. Identification with the parents' personalities varies with the nature of the feelings and attitudes of *Ss* toward their parents (parent-cathexis). Positive parent-cathexis is related to high identification scores, and negative parent-cathexis is related to low identification scores.

2. Identification with the parents' personalities varies with the extent to which the parents' personalities are congruent with the *Ss*' concepts of the ideal mother and the ideal father.

3. Cathexis for the parents varies with the extent to which the parents' traits conform with the *Ss*' concepts of ideal mother and ideal father.

4. Self-esteem varies with: (*a*) the degree of congruence between the *Ss*' traits, and their concepts of an ideal self; (*b*) the extent to which the *Ss*' traits resemble those of their parents.

Subjects

Fifty-six male undergraduates and 56 female undergraduates from psychology classes at Emory University and at the Georgia State College of Business Administration served as *Ss* in the study. Mean age was 23.69 years, *SD* 3.39, for the males, and 20.89 years, *SD* 3.39, for the females.

Materials

Ten questionnaires were constructed from a series of forty personality traits which were to be responded to by *Ss* in different ways. The traits were:

1. Sense of humor.
2. Temper.
3. Ability to express self.
4. Ability to express affection.
5. Ability to express sympathy.
6. Self-understanding.
7. Usual mood.
8. General knowledge.
9. Popularity with others.
10. Self-confidence.
11. Ability to accept criticism.
12. Sensitivity to others' feelings.
13. Intelligence level.
14. Capacity for work.
15. Ability to meet new people.
16. Self-discipline.
17. Ability to make decisions.
18. Tolerance of others' shortcomings.
19. Ability to overcome self-consciousness.
20. Ability to relax and 'let hair down.'
21. Depth of feeling.
22. Sense of responsibility.
23. Understanding of intimates.
24. Receptiveness to new ideas.
25. Attitude toward sex.
26. Ease of getting to know.
27. Personality.
28. Ability to control emotion.

[1] The author would like to express his appreciation to Miss Elizabeth Valerius, a graduate student at Emory University at the time the study was commenced; she provided invaluable assistance to the author in the preliminary stages of the work. Dr. James P. Rogers, Jr., of the Georgia State College of Business Administration helped the author in gathering part of the data.

[2] Now with the Department of Psychiatry, University of Alabama Medical Center.

29. Ability to put ideas across.
30. Degree of freedom from fear.
31. Degree of independence.
32. Ability to concentrate.
33. Ways of disciplining others.
34. Philosophy of life.
35. Religious beliefs.
36. Business sense.
37. Happiness.
38. Conformity to own moral standards.
39. Promptness in getting things done.
40. Ability to act the right way in every situation.

The traits are seen to be formulated in nontechnical language, so as to approximate the terms in which *Ss* think about and describe personality.

The questionnaires which were constructed from these traits, and the scores which were derived from them, are listed as follows:

a. Perceived similarity-to-parents questionnaire. Each *S* was requested to rate the extent to which he resembled his mother and father on each of the 40 traits, in accordance with the following instructions:

> Below are listed a number of things characteristic of yourself, or related to you. Consider each item carefully, and decide which of your parents you resemble most in this characteristic. Then, encircle the appropriate number beside each trait, in accordance with the scale which follows. If you resemble both parents on a given trait, then encircle two appropriate numbers.
>
> +3: Very close resemblance to your father in this respect
> +2: Closely similar to your father in this respect
> +1: Faintly resemble your father in this respect
> 0: Resemble neither father nor mother
> −1: Faintly resemble your mother in this respect
> −2: Closely similar to your mother in this respect
> −3: Very close resemblance to your mother in this respect

This scale appeared beside each trait on the "Similarity-to-Parents" questionnaire. Total *perceived similarity* scores were computed by summing the "plus" and "minus" numbers that had been encircled, and were designated I_f and I_m, for rated similarity to father and to mother, respectively.

b. The cathexis questionnaires. The forty traits were listed on separate forms, entitled "Feelings about Father," "Feelings about Mother," and "Feelings about the Self." The *Ss* were instructed to indicate their feelings about these traits of the respective persons by encircling a number from the five-point scale which appeared beside each trait on the separate cathexis questionnaires. The numbers signified:

1. Have strong positive feelings; like very much.
2. Have moderate positive feelings.
3. Have no feelings one way or the other.
4. Have moderate negative feelings.
5. Have strong negative feelings; dislike very much.

Total scores were obtained by summing, and were designated *FC, MC,* and *SC* for father-cathexis, mother-cathexis, and self-cathexis (self-esteem), respectively.

c. Real- and Ideal-Person ratings. Six identical trait-rating forms were prepared from the trait list. Each trait was restated in the form of a continuum, with five scale points between the extremes. For example, the item *temper* was presented as a continuum with *even tempered* at one pole, and *very quick to lose temper* at the other. The *S* encircled a number from 1 to 5, indicating where the person being rated would fall along this continuum. The socially desirable version of each trait was randomly ordered throughout the list, so that for some traits, a score of 1 would indicate the desirable pole, while for others, a score of 3 or 5 would represent the desirable degree. This randomization was done in order to discourage response sets.[3]

The rating forms which the *Ss* completed were entitled *Real Father, Ideal Father, Real Mother, Ideal Mother, Real Self,* and *Ideal Self.* The *S* was instructed to "Put a circle around the number which best describes the person, with respect to each trait."

Two sets of discrepancy scores were calculated from these rating scales. One set pertained to the discrepancy between "real" and "ideal" ratings for mother, father, and the

[3] One of the indicators of response sets is a marked preference on the part of *Ss* for some one response category, e.g., for high, middle, or low numbers on a numerical rating scale. Inspection of the data obtained in the present study showed considerable intraindividual variability, from questionnaire to questionnaire, in responding to items. While this is not proof that response sets were eliminated, it supports the contention that they were minimized.

self. These were obtained by subtracting the "real" ratings from the "ideal" ratings, ignoring sign, and then summing. Total "real-ideal" discrepancy scores were designated *Dis-*F_i-F_r, *Dis-*M_i-M_r, and *Dis-*S_i-S_r.

The other set of discrepancy scores computed from these forms provided a further index of similarity to parents, which is called the *derived similarity score*. These scores were obtained by comparing "Real Father" ratings with "Real Self" ratings, and "Real Mother" ratings with "Real Self" ratings, ignoring sign. Each pair of scores yielded a discrepancy with a possible range from 0 to 4. These discrepancies were summed, and designated *Dis-*F_r-S_r and *Dis-*M_r-S_r for differences between father's personality and the self, and mother's personality and the self, respectively.

In summary, each *S* completed a total of ten questionnaires. The names of these questionnaires, the order in which they were presented to the *S*s, and the scores which were obtained from them are as follows:

(*a*) Similarity to parents (I_f and I_m—*Perceived similarity scores*)
(*b*) Feelings about father's personality (*FC*—*Father-cathexis scores*)
(*c*) Feelings about mother's personality (*MC*—*Mother-cathexis scores*)
(*d*) Feelings about the self (*SC*—*Self-cathexis scores*, the measure of self-esteem)
(*e*) Ideal Mother ratings
(*f*) Real Self ratings
(*g*) Real Father ratings
(*h*) Ideal Self ratings
(*i*) Real Mother ratings
(*j*) Ideal Father ratings

Discrepancies were computed between each pair of traits and summed through all traits, on forms (*e*) and (*i*) [*Dis-*M_i-M_r], (*g*) and (*j*) [*Dis-*F_i-F_r], and (*f*) and (*h*) [*Dis-*S_i-S_r]. These are the discrepancy-between-"real"-and-"ideal" ratings of the mother, father and self, and are termed *congruence measures* throughout the tables. Further discrepancies were determined and summed between (*f*) and (*g*) [*Dis-*F_r-S_r] and between (*f*) and (*i*) [*Dis-*M_r-M_r]. These latter comprise the two additional measures of similarity between the self and the parents, and are termed *derived similarity scores*.[4]

Table 1

Means, Standard Deviations, and Corrected Reliability Coefficients of the Scales

Scale	Men			Women		
	Mean	*SD*	*r*	Mean	*SD*	*r*
I_f	38.35	22.02	.91	38.31	20.36	.88
I_m	34.65	20.65	.90	37.82	19.07	.80
Dis-F_r-S_r	34.89	15.59	.84	37.62	14.70	.88
Dis-M_r-S_r	33.95	17.25	.88	35.09	12.67	.86
FC	86.02	26.18	.92	81.40	26.24	.94
MC	89.40	24.23	.92	87.95	23.39	.94
SC	87.80	22.23	.92	90.55	19.76	.90
Dis-F_i-F_r	36.57	21.40	.93	31.43	21.00	.93
Dis-M_i-M_r	35.76	18.29	.89	35.95	18.71	.92
Dis-S_i-S_r	37.18	19.69	.92	39.96	14.53	.86

Testing Procedure

The questionnaires were stapled as a booklet and distributed to the *S*s during a classroom period. Instructions for filling them out were presented orally. Date of birth, sex, and marital status were the only identifying information that was requested from the *S*s, on the premise that relative anonymity would encourage frankness in response. The *S*s took the booklets home with them, and returned the completed forms on a following day.

Results

General Results and Questionnaire Reliability

Means, standard deviations, and reliability coefficients for all variables are shown in Table 1. Each scale has satisfactory reliability.

None of the mean scores obtained by males and females on each scale showed any significant differences.

Identification with Parents and Parent-Cathexis

Hypothesis 1 may be restated: "If you *like* your parents, you tend to *be like* them."

[4] The two measures of similarity-to-parents—perceived similarity scores and derived similarity scores—did not correlate very highly with each other. I_f correlated with *Dis-*F_r-S_r only −.35 for males, and −.20 for females; I_m correlated with *Dis-*M_r-S_r only −.17 among the males, and −.43 among the female *S*s. In spite of their near-independence of each other, the perceived- and derived-similarity scores correlated in predicted ways with other variables.

Table 2

Correlations Between Similarity-to-Parents Scores and Parent-Cathexis Scores

Index of Similarity	Men		Women	
	FC	*MC*	*FC*	*MC*
I_f	−.52**		−.45**	
Dis-F_r*-*S_r	.43**		.28*	
I_m		−.38**		−.62**
Dis-M_r*-*S_r		.20		.38**

* Significant at .05 level.
** Significant at .01 level.

Table 2 shows that both the perceived (I_f and I_m) and the derived (*Dis-*F_r*-*S_r and *Dis-*M_r*-*S_r) indices of identification, or similarity to the parents, are significantly correlated with the parent-cathexis measures. In one instance only did an *r* fall short of significance, namely, that between *Dis-*M_r*-*S_r and *MC* for males. Transformation of the *r*'s in this table (and all others) to *z*'s was performed, in order to permit cross-sex comparisons of the magnitude of *r*. None of the observed sex differences between *r*'s in Table 2 were significant.

Identification with Parents and Parental Congruence with Ideals

The second hypothesis predicted that *Ss* tend to "identify" only with those parental traits that coincide with their ideals for those traits. Thus, the larger the discrepancy between the actual (rated) traits of the parents, and the *Ss*' concepts of ideal parent, then the *smaller* are the *perceived similarity* scores (I_f and I_m), and the *larger* are the *derived similarity* scores. Table 3 shows that the obtained *r*'s among the relevant variables support the predictions in magnitude and in direction. It may be concluded that *Ss* tend to identify only with those parental traits which are exemplary for them; i.e., which coincide with their ideals.

Only one significant sex difference in *r*'s was obtained in Table 3; I_m correlated − .26 with *Dis-*M_i*M*$_r$ for males, and − .62 for females. This difference yielded a *t* ratio for the corresponding *z*'s of 2.37 (*p* about .02). Evidently among males, perceived similarity with the mothers' traits is more independent of the mothers' congruence with ideals than is the case for female *Ss*.

Parent-Cathexis and Parental Congruence with Ideals

It was predicted that the cathexis response to parents' traits varies with the magnitude of the discrepancy between the actual (rated) traits of the parents, and the *Ss*' ideals for parental traits. Table 4 shows that there is a substantial correlation between the cathexis measures and the extent of congruence between parents' traits and the *Ss*' ideals for those traits. It is further noted that *MC* is more highly related to *Dis-*M_i*-*M_r among female *Ss* (.75) than among males (.29). The *t* ratio for the difference between the corresponding *z*'s was 3.46 (*p* less than .01). This difference suggests that among males, mother-cathexis is more independent of the mother's conformity with concepts of the ideal-mother than is the case among females.

Self-Esteem, Identification, and Congruence of Ideal Self with Real Self

Self-cathexis (*SC*) scores indicate the extent to which a person likes or dislikes his own traits. The *SC* scores may thus be regarded as indices of self-esteem. As with the parent-cathexis scores, *SC* correlated with the relevant discrepancy scores (*Dis-*S_i*-*S_r) with significances beyond the .01 level (.62 for males, and .53 for females).

The fact of covariance between *SC* scores and the identification estimates (I_f, I_m, *Dis-*F_r*-*S_r and *Dis-*M_r*-*S_r) is shown in Table 5. The index of self-esteem (*SC*) is significantly

Table 3

Correlations Between Similarity-to-Parent Scores and Parental Congruence with Ideals

Index of Similarity	Men		Women	
	Dis-F_i*-*F_r	*Dis-*M_i*-*M_r	*Dis-*F_i*-*F_r	*Dis-*M_i*-*M_r
I_f	−.57**		−.44**	
Dis-F_r*-*S_r	.44**		.36**	
I_m		−.26*		−.62**
Dis-M_r*-*S_r		.64**		.44**

* Significant at .05 level.
** Significant at .01 level.

correlated with all measures of similarity to the parents among male *Ss*, and with all indices among the females except *Dis-F_r-S_r*. The relative slight magnitude of the *r*'s suggests that self-esteem is only partly determined by parental identification, probably because identification itself seems to occur only when the parents' personalities are exemplary.

Discussion

General Results

The fact that none of the comparisons between mean scores of males and females that were reported in Table 1 showed significant differences warrants some comment. One might expect, for example, that males would obtain higher similarity-to-father scores than women, and that women would obtain higher similarity-to-mother scores than men. Further, it might be predicted that men would show greater similarity to their fathers than to their mothers, and vice versa for the female *Ss*. The lack of significant differences on these variables may possibly be attributed to the fact that the traits employed in the questionnaires are highly general in nature, and hence cannot be readily sex-typed.

Identification with Parents, Parent-Cathexis, and Parental Congruence with Ideals

Perceived- and derived-similarity scores were shown to be correlated significantly with the parent-cathexis scores. These findings lend support to Mowrer's theory of "developmental identification" which, loosely restated, asserts that positive feeling toward the parents fosters identification with their traits (6). It should be noted that the similarity scores used herein are at best only crude indices of identification.

Table 4

Correlations Between Parent-Cathexis Scores and Parental Congruence with Ideals

Congruence Measure	Men		Women	
	FC	*MC*	*FC*	*MC*
Dis-F_i-F_r	.83**		.78**	
Dis-M_i-M_r		.29*		.75**

* Significant at .05 level.
** Significant at .01 level.

Table 5

Correlations Between Self-Cathexis Scores and Measures of Similarity to the Parents

Index of Similarity	Men	Women
I_f	−.37**	−.43**
Dis-F_r-S_r	.50**	.19
I_m	−.38**	−.32*
Dis-M_r-S_r	.51**	.28*

* Significant at .05 level.
** Significant at .01 level.

It is always possible that the *Ss*' observations and judgments of similarity were *autistic*, i.e., strongly influenced by their feelings toward their parents. Indeed, the very fact of correlation between cathexis scores and similarity scores suggests that autism may have influenced their judgments. Finally, any real similarity between the *Ss* and their parents may have derived from sources other than identification, viz., patterned socialization and "direct tuition" practises (1, pp. 56–59) which produced similarities "by accident."

The findings with respect to identification and parental congruence with ideals suggest that identification is a *selective* process. It may be hypothesized that children do not identify with *all* of their parents' traits; rather, they select those traits which seem *worthy* of emulation, which will be instrumental in the attainment of assorted valued ends, viz., parental approval, success in achievements, etc. (cf. 3, 6).

Parent-Cathexis and Parental Congruence with Ideals

Cathexis has been redefined from its original psychoanalytic meaning by Parsons. He states that cathexis refers to "a state of the organism—a state of euphoria or dysphoria—*in relationship to some object*. . . . It is *object-oriented affect*. . . . It involves attaching affective significance to an object" (8, p. 10; his italics). In the present context, parent-cathexis refers to the feelings of liking-disliking for parents' traits. The idea that cathexis-responses might be related to the degree of congruence between actual or perceived characteristics of the object, and ideals pertaining

to that object, was suggested by an earlier work dealing with body-cathexis (4). As a hypothesis for further exploration, it is suggested that cathexis for other persons, the self, or any object is in part a function of the congruence of the object with *Ss*' ideals. This hypothesis has bearing on the theory of interpersonal attractiveness (7), sociometry, aesthetics, etc.

Self-Esteem, Identification, and Congruence between Ideal Self and Real Self

Self-esteem—positive cathexis for the self—appears to hinge upon congruence between the "real self" and the "self-ideal." The obtained correlations between *SC* scores and similarity scores suggest that the parents' personalities may have served, not only as a model for the "real self" of the *Ss*, but also as the model for their self-ideals. Most psychoanalysts, in fact, trace the origin of the self-ideal (superego and ego-ideal) to identification with parents. The present findings may be viewed as partial support to the psychoanalytic theory of the relationship between identification and the superego (2, pp. 103–105).

Summary

Fifty-six male and 56 female college students were tested with a series of questionnaires designed to measure similarity with parents' personalities, parent-cathexis, and self-cathexis.

Scores indicative of the degree of similarity between the self and the parents were found to vary with parent-cathexis, and with the degree of congruence between the parents' personalities and the *Ss*' concepts of the ideal parent.

Parent-cathexis scores were found to vary with the degree of congruence between the parents' rated personalities, and *Ss*' concepts of the ideal parent.

Self-cathexis (a measure of *self-esteem*) was found to vary with (*a*) the congruence of the "real self" with the "self-ideal" and (*b*) with the degree to which *Ss* resembled their parents' personalities.

The findings were related to the theory of identification, and to a general theory of the cathexis-response to objects.

The overall results appear to confirm the clinical observation that identification with parents, feelings and attitudes to parents, and self-esteem all cohere as a syndrome (cf. 5, pp. 31–62).

Received November 21, 1956.

References

1. Cameron, N., & Magaret, Ann. *Behavior pathology*. Boston: Houghton Mifflin, 1951.
2. Fenichel, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
3. Hendrick, I. Early development of the ego: Identification in infancy. *Psychoanal. Quart.*, 1951, **20**, 44–61.
4. Jourard, S. M., & Secord, P. F. Body-cathexis and the ideal female figure. *J. abnorm. soc. Psychol.*, 1955, **50**, 243–246.
5. Maslow, A. H. *Motivation and personality*. New York: Harper, 1954.
6. Mowrer, O. H. *Learning theory and personality dynamics*. New York: Ronald Press, 1950.
7. Newcomb, T. The prediction of interpersonal attraction. *Amer. Psychologist*, 1956, **11**, 575–586.
8. Parsons, T., & Shils, E. A. (Eds.) *Toward a general theory of action*. Cambridge, Mass.: Harvard Univer. Press, 1951.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

Perceptions of Significant Family and Environmental Relationships in Aggressive and Withdrawn Children[1]

David S. Shapiro[2]
Teachers College, Columbia University

The mediating function of cognitive processes has been given significant attention in a variety of theories and experiments. Tolman (11), for example, emphasizes the importance of expectancies in guiding behavior; Hull (4) has suggested that pure stimulus acts contribute stimuli necessary for problem solving; Mowrer (7) regards preparatory set as a determinant of behavior. That such factors as value, need, and expectancy may influence perceptions has been demonstrated in experiments reported by McClelland and Liberman (6), Vanderplas and Blake (12), Bruner and Postman (1), and others (2, 3, 5, 8, 9).

The present study is an attempt to extend investigations of the relationship between cognitive and other behavioral processes to the problem of personal adjustment. The specific question dealt with was whether characteristic patterns of overt adjustment are systematically associated with characteristic forms of perceptual behavior. Pursuit of this inquiry entailed the identification of groups of subjects differing markedly in overt adjustment patterns, the formulation of hypotheses accounting for such behavioral differences in terms of differing cognitions, and a final determination of relationships between the cognitive and adjustment variables.

Aggression and withdrawal in school-age children were selected as the adjustment variables. Using as a guide the phenomenological approach (10), which postulates a one-to-one relationship between an individual's behavior and his psychological field, two factors in the perceptual fields of aggressive and withdrawn children may be considered significant in determining behavior. The first factor is the degree to which the individual perceives the world to be threatening. Thus, aggressive children, apparently less deterred by threat, are likely to perceive the world as less threatening than do withdrawn children. It is equally possible that aggressive and withdrawn individuals do not differ in their perceptions of threat but that, in the face of equal dangers, a second factor, the individual's perception of his resources for dealing with the threat, becomes the significant cue determining overt behavior. It was predicted, therefore, that aggressive and withdrawn children would differ in the degree to which they perceive the world to be threatening and in their perceptions of their resources for dealing with threat. Specifically, the hypotheses tested in this study are that in perceptual tasks related to significant family experiences, aggressive children:

1. Estimate the size, strength, and ability of child figures to be greater than do withdrawn children and, similarly, estimate the size, strength, and ability of child figures to be greater in relation to those of the parental figures than do withdrawn children.

2. Expect punishment for various acts to be less severe than do withdrawn children.

3. Describe the immediate outcome of situations as more favorable than do withdrawn children.

Subjects

The subjects used in this study were 101 boys aged six through ten selected from among 1,531 children in the first five grades of a public school system in a small New York

[1] Based upon a doctoral dissertation submitted to Columbia University in 1956. The author is indebted to Professors E. J. Shoben, Jr., L. F. Shaffer, and R. L. Thorndike for their assistance.

[2] Currently at the Bradley Center, Inc., Columbus, Georgia.

State city. Selections were made by a committee consisting of a director of elementary guidance, a school psychologist, school principals, and classroom teachers. The subjects were divided into an experimental group of 33 boys manifesting persistently aggressive behavior (Group A); an experimental group of 35 boys manifesting persistently withdrawn behavior (Group W); and a control group of 33 well-adjusted boys (Group N). The groups were found to be generally similar in age, grade, and social class membership distributions. Grade achievement scores on reading and arithmetic tests, used as an estimate of intelligence, showed the aggressive and withdrawn groups to be generally similar. The well-adjusted group scored higher on both tests than did the aggressive or the withdrawn groups. These results, significant at the 5 per cent level of confidence, may be attributed, in part, to the clinically observed tendency for disturbed children to be less efficient in learning. It is also probable, however, that moderately brighter children were selected as members of the well-adjusted group.

Experimental Tasks

Each subject was given two perceptual tasks: (*a*) to answer questions about seven pictures of a family group in presumably anxiety-arousing situations (the Family Situations Task); and (*b*) to draw a picture of his parents and himself (the Family Drawings Task). These individually administered tasks were designed to assess the subject's perceptions of five variables pertinent to the hypotheses being tested. The variables were:

1. Size, measured by the height in centimeters of drawings of parent and child.
2. Strength, measured by responses to questions concerning tests of physical strength between a father and son.
3. Ability, measured by descriptions of the behavior of parent and child figures in coping with critical situations.
4. Expectations regarding punishments.
5. Expectations regarding the outcomes of situations.

The Family Situations Task consisted of seven original pictures of a family group which were drawn by an artist from detailed instructions. A questionnaire containing thirty items was developed to accompany the pictures. Photographic reproductions of the pictures were used in administering the task. Brief descriptions of the pictures are presented:

Card One. Father, mother, and a little boy are walking into the dining room to have breakfast. Their dog is pulling at the tablecloth with all the food on it and threatening to spoil the breakfast.

Card Two. The father and son are having a weight-lifting contest. Seven differently sized weights are in front of them.

Card Three. The family is reading in the living room. They are so engrossed in their reading that they do not notice a fire that has started in a curtain.

Card Four. An obviously angry father is confronting a boy.

Card Five. The parents are sitting at a table in the kitchen, engrossed in conversation. The boy is running in to them, obviously excited.

Card Six. Father and son are having a tug-of-war.

Card Seven. The family is standing in front of a lion's cage at the zoo. The lion is angry and has bent a bar of his cage.

The Family Drawings Task consisted of a single drawing made by the subject of his parents and himself. The directions were: "Make a fast drawing of your father, your mother, and you. Make everybody standing up." Drawings were done on a single sheet of white paper, 8½ by 11 inches.

Procedures in the Analysis of Data

Data for the size and strength variables of the first hypothesis were scored objectively. Size was measured by the height in centimeters of drawings of parents and child figures; strength was measured by verbal responses to seven questions relating to tests of physical strength between a father and son. The third variable dealt with in the first hypothesis, ability, was measured by ratings of verbal descriptions of the behavior of parent and child figures in coping with situations involving varying degrees of threat to comfort and safety. Ability was rated by three judges on a four-point scale of ability to cope effectively with the threatening aspects of the situation.

The second hypothesis, which contended that aggressive children expect punishment for various acts to be less severe than do withdrawn children, was measured entirely by ratings of the subject's description of punishment inflicted upon a boy for various offenses.

Table 1

Heights of Family Figures and Ratio of Height of Boy to Height of Father (Measured in centimeters)

Figure	Group A		Group W		Group N		Difference Between Groups	
							A vs. W	A & W vs. N
	M	*SD*	M	*SD*	M	*SD*	*t*	*t*
Boy	7.04	3.56	6.18	2.96	5.36	2.09	1.08	2.29*
Father	9.31	3.92	8.76	3.59	7.57	2.54	.64	2.28*
Mother	9.29	4.08	8.43	3.97	7.08	2.44	.88	2.72*
Boy/Father (ratio)	.77	.25	.72	.23	.72	.19	.86	.11

* Significant at .05 level.

Three judges rated the responses on a four-point scale of severity of punishment.

The third hypothesis, asserting that aggressive children describe the immediate outcomes of situations more favorably than do withdrawn children, was measured by responses to seven questions requiring the subject to predict the outcome of situations. Five of the questions were objectively scored, while two questions were rated on a four-point scale of severity or mildness of outcome. In all cases where ratings were used, agreement of the judges on 85 per cent of the ratings was set as a minimum standard of reliability.

Results

The results obtained for the first hypothesis indicated that aggressive children do tend to perceive child figures to be stronger in relation to parent figures than do withdrawn children, but that the two groups have essentially similar perceptions regarding size and ability. The hypothesis that aggressive children expect punishment to be less severe than do withdrawn children was not supported. Substantial evidence was found, however, for the hypothesis that aggressive children describe the outcomes of situations as more favorable than withdrawn children.

Size, Strength, and Ability

Table 1 shows that the disturbed children (Group A and Group W) tended to have similar perceptions of the size of child figures and parent figures. On the other hand, Table 1 demonstrates that there was a difference between the disturbed children, taken as one group, and the nondisturbed children (Group N), with regard to the size variable. The disturbed children drew both parent and child figures taller than did the nondisturbed children. The disturbed and the nondisturbed groups did not differ, however, with respect to the ratio between the height of the boy figures and the height of the father figures. The finding of absolute size differences without associated differences in the perceptions of relative size has only minor relevance, since the absolute size differences cannot be explained in terms of differing perceptions of threat or resources for dealing with threat.

In their responses to two of the seven questions concerning relative strength, the aggres-

Table 2

Estimated Weight-Lifting Ability in Pounds

Situation	Group A	Group W	Group N	Differences Between Groups
				A vs. W
Boy's present ability				
Mean	61.0	40.9	50.7	
SD	30.8	25.9	19.7	
t				2.86**
Boy's future ability				
Mean	161.4	117.9	126.8	
SD	84.1	46.4	48.0	
t				2.58**

** Significant at .01 level.

Table 3

Optimistic versus Pessimistic Outcomes of Situations

Situation	Groups Compared: A vs. W χ^2	A vs. W p	A & W vs. N χ^2	A & W vs. N p
Save or spoil breakfast?	5.81	<.02*		
Extent of damage from fire?	.02	<.99	4.04	<.30
Will parents give boy money?	8.43	<.01**		
Father lost his job, what will happen to the family?	4.61	<.30	.07	.99
Long, short, or medium amount of time for father to find another job?	4.12	<.20	1.43	<.70
Amount of time needed to find another job?	2.70	<.30	3.14	<.30
What will happen to family if lion breaks out of cage?	1.45	<.70	4.03	<.30

* Significant at .05 level.
** Significant at .01 level.

sive children perceived child figures to be stronger in relation to parental figures than did the withdrawn children; the responses to the other five questions showed no differences between the experimental groups. Table 2 presents the results for the two questions for which significant findings were obtained.

Statistical analysis of the results obtained for the ability variable, using 24 chi-square tests, produced no statistically significant findings. The reliability of the results obtained for this variable may be questioned, however, for several reasons. Two of the three pictures used to obtain measures of relative ability, the fire scene and the lion scene, deal with situations which not all of the children could be expected to relate to their own experiences. In addition, the questions used for this variable permitted wide latitude in responses; such questions were consistently unrevealing of differences between groups throughout the study.

Punishment

The results obtained for the punishment variable do not support the hypothesis that aggressive children expect punishment to be less severe than do withdrawn children. The statistical analysis of the ratings of the verbal responses, using six chi-square tests, produced no results indicative of differences between the experimental groups or between the disturbed and the nondisturbed children. It appears, therefore, that aggressive and withdrawn children, despite marked differences in overt behavior, tend to have similar expectations regarding the severity of punishment. It should be noted, however, that only one picture was used in measuring this variable and that the questions used permitted wide latitude in responses.

Expected Outcomes

The findings concerning expected outcomes, as shown in Tables 3 and 4, support the hypothesis that aggressive children describe the immediate outcomes of situations as more favorable than do withdrawn children. Table 3 demonstrates that significant affirmative results were obtained for two of the seven questions. When responses to all seven questions were dichotomized as tending to be either

Table 4

Number of Situations Described as Having Favorable Outcomes[a]

Measure	Group A	Group W	Group N	Differences Between Groups: A vs. W
Mean	3.49	2.23	3.42	
SD	1.31	1.46	1.63	
t				3.69**

[a] An optimism score was arrived at for each subject by determining how many of his responses to the seven questions of this variable tended to predict favorable outcomes.
** Significant at the .01 level.

optimistic or pessimistic, the mean number of optimistic responses for aggressive children was found to be significantly greater than that for withdrawn children, as shown in Table 4. These findings constitute substantial evidence for the validity of this hypothesis.

General Conclusions

The results obtained support the conclusion that differences in the overt adjustment patterns of aggressive and withdrawn children are related to differences in expectations about the outcomes of situations that are relevant to the needs and security of the individual. It would appear that aggressive children are guided in their overt behavior by a greater degree of confidence in a favorable outcome than are withdrawn children. The evidence that aggressive children tend to perceive child figures to be stronger in relation to parent figures than do withdrawn children is considered to be most economically explainable as further support for the conclusion that aggressive children tend to be more optimistic than do withdrawn children.

Discussion

The results are consistent with the general contention that behavior is mediated by central, cognitive processes. The factor of mood, with aggressive children tending to be optimistic, and withdrawn children tending to be pessimistic, emerges as significant in guiding the behavioral adjustments of children. Predictions that aggressive and withdrawn children differ in the degree to which they perceive the world to be threatening, and in their perceptions of the adequacy of their resources for dealing with threat were, in general, not supported. In view of the over-all results of the study, the differences in the overt adjustment patterns of aggressive and withdrawn children appear to be most adequately and simply explainable in terms of reward learning theories. Aggressive children appear to have learned from experience that aggression may lead to reward, either in the attainment of goals or in the reduction of anxiety, while withdrawn children, perhaps having had less fortunate experiences in meeting needs in a direct manner, appear to have learned that withdrawal and renunciation of goals are most effective in allaying anxiety.

Summary

This study sought to determine whether characteristic patterns of overt adjustments are systematically associated with characteristic forms of perceptual behavior. Specifically, it was hypothesized that in perceptual tasks related to significant family experiences, aggressive children estimate the size, strength, and ability of child figures to be greater both in absolute terms and greater in relation to the size, strength, and ability of parental figures than withdrawn children; that they expect punishment to be less severe, and describe immediate outcomes of situations as more favorable than withdrawn children. The results lead to the conclusion that differences in the overt adjustment patterns of aggressive and withdrawn children are related to differences in expectations about the outcomes of situations relevant to the needs and security of the individual.

Received December 10, 1956.

References

1. Bruner, J. S., & Postman, L. Perception, cognition, and behavior. *J. Pers.*, 1949, **18**, 14–31.
2. Cowen, E. L., & Beier, E. G. The influence of "threat-expectancy" on perception. *J. Pers.*, 1950, **19**, 85–94.
3. Diven, K. Certain determinants in the conditioning of anxiety reactions. *J. Psychol.*, 1937, **3**, 291–308.
4. Hull, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
5. Lazarus, R. S., Yousem, H., & Arenberg, D. Hunger and perception. *J. Pers.*, 1953, **21**, 312–328.
6. McClelland, D. C., & Liberman, A. M. The effect of need for achievement on recognition of need-related words. *J. Pers.*, 1949, **18**, 236–251.
7. Mowrer, O. H. *Learning theory and personality dynamics*. New York: Ronald Press, 1950.
8. Razran, G. H. S. A quantitative study of meaning by a conditioned salivary technique (semantic conditioning). *Science*, 1939, **90**, 89–90.
9. Riess, B. F. Semantic conditioning involving the galvanic skin reflex. *J. exp. Psychol.*, 1940, **26**, 238–240.
10. Snygg, D., & Combs, A. W. *Individual behavior*. New York: Harper, 1949.
11. Tolman, E. C. *Purposive behavior in animals and men*. New York: Appleton-Century, 1932.
12. Vanderplas, J. N., & Blake, R. R. Selective sensitization in auditory perception. *J. Pers.*, 1949, **18**, 252–266.
13. Warner, W. L., Meeker, M., & Eells, K. *Social class in America*. Chicago: Science Research Associates, 1949.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

The Role-Taking Hypothesis in Delinquency[1]

Charles F. Reed
Upstate Medical Center, State University of New York
and Carlos A. Cuadra
RAND Corporation

On the basis of an hypothesis concerning psychopathy, Gough has devised a scale for the identification of predisposition for delinquent behavior (2, 6, 7). The present study is an experimental examination of the operation of the hypothesized variable in the production of scores on the scale.

Gough finds the psychopath unable to "look upon the self as a 'social object'" or to "elaborate an adequate and realistic set of social expectancies and critiques" (7, p. 207). For some clinical observers, however, one of the chief characteristics of psychopaths is their singular facility for slipping from one role to another, assuming the demeanor most opportune for the moment. Gough emphasizes that it is not a deficiency of role-playing ability to which he refers, but an absence of the "residuals which ordinarily accrue as a consequence of interactional experience . . ." (7, p. 207). However, these residuals are not identified specifically, and Gough's concern seems to remain with the supposed inability to see one's self as seen by others.

Gough's Delinquency Scale (*De*)[2] incorporates this hypothesis and appears to have practical screening efficiency in differentiating between delinquent and nondelinquent samples (5, 6, 7). The item "I often think about how I look and what impression I am making upon others" (answered False) is given as an example of a cluster of items reflecting "role-taking deficiencies, insensitivity to interactional cues and the effects of one's own behavior on others" (7, p. 209). Items such as this seem rather to be reports of indifference to the opinion of others, not direct manifestations of incapacity in role-taking. Whatever the basis for the efficacy of *De* for screening, the data presented by Gough do not warrant the conclusion that incapacity in role-taking has been demonstrated. It could be argued with equal justification that part of the scale's power lay in items admitting past delinquencies, e.g., "I have never been in trouble with the law" answered False, since seven items in the scale are of this nature.

Data pertaining to the relative social sensitivity of subjects who receive high scores on *De* will be presented in this report. Social sensitivity was defined experimentally as the ability of a subject to predict descriptions made of him by peers. A normal sample was chosen instead of a sample of delinquent subjects for methodological reasons concerned with equivalence of task and with acquaintanceship variables. The study deals with the examination of a delinquency scale rather than with delinquents themselves. References to "High *De* scorers" are not intended to convey the impression that these subjects were delinquent.

Procedure

Successive classes of student nurses undergoing training at a VA neuropsychiatric hospital were used as *Ss*. Each class was tested

[1] This study was inaugurated while both investigators were members of the staff of the Clinical Psychology Service of the VA Hospital, Downey, Illinois. The authors wish to acknowledge the support and assistance of the hospital administration and the Nursing Education Service throughout the course of the project.

[2] Now renamed the Socialization (*So*) Scale and scored in reverse in the California Psychological Inventory (6).

near the close of three months' residence at the hospital, where the students had worked on wards, attended classes, and lived together. There was ample opportunity for observation of each other in a variety of working conditions, some of them stressful.

For the purposes of this study, each *S* was assigned to a group of four *S*s. Each *S* was asked to describe herself on an adjective checklist (I), marking every adjective in the list with a plus when it was true or generally true, with a minus when it was false or generally false. As a second task (II), the *S* described each of the other members of her group of four. Finally, she attempted to predict how she would be described by the members of the small group as a group (III). Precautions of anonymity and seating were taken to foster candid appraisals.

The forced-choice technique was used because, in a pilot study with Navy hospital corps students in which the adjectives were checked only when they applied to a subject, there seemed to be a spurious accuracy in predicting group description. The fewer adjectives chosen, the higher the accuracy score. By employing forced choices, the correlation between accuracy and the number of adjectives marked plus was reduced to zero.

The California Psychological Inventory, of which the scale is a part, was administered as a final procedure. Complete data were available for 204 *S*s; partial data existed for 25 more *S*s.

Quantitative Data

The correspondence between an *S*'s predicted description (III) and the descriptions which are made of the *S* by her peers (composite II's) is the variable of chief interest. A score was given whenever, on a given adjective, two of the three peers recorded the same sign as that recorded by the *S*. Total score, Predictive Accuracy, is the sum of the adjectives on which such agreement occurred.

The correspondence between *S*'s self-description (I) and the peer-description (composite II's) was scored in a similar manner and the total score called Self-peer Correspondence.

Additional scores were evolved in the course of analysis of the data. They will be described after the report of initial results.

Peer-nomination Data

A second major source of data was provided by material gathered some months after the initial part of the study was completed. All *S*s were contacted by mail and asked to select from a list of their classmates five whom they considered the most insightful of the class and five whom they considered least insightful. In-

Table 1

Comparison of Scores of High-*De* Group and Low-*De* Group

Variable	High-*De* group		Low-*De* group		CR
	Mean	*SD*	Mean	*SD*	
Predictive Accuracy (III vs. II)	161.08	20.45	178.38	13.81	4.90**
Self-peer Correspondence (I vs. composite II's)	161.56	17.70	173.21	15.66	3.44**
Conventional-word Count (on III)	356.73	199.62	519.46	133.91	3.41**
Conventional-word Count (on I)	382.95	148.36	509.50	119.50	3.25**
Stereotypy	126.12	32.46	148.39	25.08	3.79**
Favorable Self-description	20.58	7.83	26.45	6.71	3.99**
Favorable Peer-description	29.52	5.41	30.15	5.19	.58
Favorable Expected Peer-description	19.24	9.20	26.09	6.00	4.36**
Anticipated Disagreement	47.40	16.95	37.32	14.07	3.20**

** $P < .01$.

sight was defined as follows: "An insightful person has the ability to recognize and understand the motives underlying her behavior and is aware of the effects of her behavior on other persons. She is alert to what other people think of her as a person. (In making judgments, do not be influenced by intelligence, likeability, etc., which are not necessarily related to insight. An unpleasant person, for example, still could be an insightful person.)" Approximately 60 per cent of the rating forms were returned.

Results and Discussion

Comparison of the scores of *Ss* at the extreme quartiles of the *De* distribution yielded the results summarized in Table 1.

Accuracy Scores

The finding of chief interest is the relative inaccuracy of the high *De* scorers in predicting how they will be described by their peers. The difference is significant beyond the .001 level. Even if the highest quartile is compared with the remainder of the sample in a one-tailed test the difference between means on Predictive Accuracy yields a critical ratio of 2.72, significant at .003. The correlation between Predictive Accuracy and *De* for the entire sample is − .41.

Several factors which might attenuate the inference to be made from this finding were examined.

The task of describing another person may be restricted by certain unverbalized conventions. Some characteristics may be assumed by most members of a group to be properties of most other members of the group and of themselves. In the present sample, for example, the adjective "intelligent" was generously used. A spuriously high accuracy score may be obtained by the *S* who is not responding to the particular members of her group of four, but who is accurately gauging the probability of an adjective being checked plus or minus when any *S* was describing any other *S*.

In this regard, it is interesting that the Self-peer Correspondence score also shows the high *De* scorers differing significantly from the low scorers. Correlation between the two scores for the entire sample is − .29. Apparently, the difference between the extreme groups is not solely a matter of anticipating the verbal responses of peers. The high *De*'s do not just mis-guess the descriptions that will be made of them. Their "private" self-descriptions differ more from the descriptions made by their peers than do the self-descriptions of the low *De*'s from their peer-descriptions.

To investigate further the possible effects of the popularity or unpopularity of adjectives, a Conventional-word Count was made. Each adjective was ranked according to frequency of use in the peer-descriptions (Checklists of Task II) and assigned a corresponding scoring weight. Self-descriptions (I) and predicted-descriptions (III) were scored for the sum of the weighted scores of adjectives marked with a plus. A high score indicates that *S* marked adjectives in the direction which conformed to the group's conventional usage.

High *De*'s marked adjectives less conventionally, both in the prediction task and in describing themselves. Conventional-word Count on Task III correlates .35 with Predictive Accuracy for the total sample. When Conventional-word Count is held constant, the correlation between *De* and Predictive Accuracy drops from − .41 to − .31, a coefficient still significantly different from zero at the .01 level. Conventionality of description—if it is to be construed as an attenuating factor—still fails to delete the differences in accuracy between the *De* extremes.

Stereotypy is a score based on the number of adjectives which an *S* checked in the same direction in describing all three peers in her group. A high score would seem to indicate either a failure to discriminate or a highly homogeneous group. High *De* scorers earn significantly lower scores. This score may simply be another reflection of venturing from popular responses.

The adjective checklists were scored for use of favorable and unfavorable items on the basis of the ratings procured from a group of judges by Gough (4). Favorable Self-description refers to the relative emphasis on favorable adjectives in describing self. Favorable Expected Peer-description refers to the relative emphasis on favorable adjectives in Task III; Favorable Peer-description indicates the

relative representation of favorable adjectives ascribed to *S* by her peers.

The results summarized in Table 1 indicate that while high *De* scorers described themselves in less favorable terms and expected to be described in less favorable terms than did the low scorers, the descriptions which were made of them were not less favorable.

A final score, Anticipated Disagreement, is the sum of adjectives given opposite signs by an *S* in I and in III. A high score may be interpreted as indicative of an expectation that the group will differ from the *S*'s self-description. High *De* scorers expect to be misdescribed or perhaps misunderstood. This is not the same phenomenon of describing self in unfavorable terms and expecting to be described in unfavorable terms, since this score is based on adjectives which differ in sign between self- and expected-description. This apparent psychological isolation may be related to characteristics Gough found in high scorers: "feelings of despondency and alienation, lack of confidence in self and others" (7, p. 209).

Peer-nomination Findings

Nominations for insightfulness were tallied for each *S* and converted into *T* scores. The ratings for insight for the *S*s in extreme quartiles on the *De* distribution showed a difference significant beyond the .02 level, *S*s with high *De* scores being rated less insightful.

Content of Adjectives

The following adjectives are associated with scores in the extreme quartiles of the *De* distribution of the sample. Adjectives discriminating at the .01 level are listed before the ellipsis, those at the .05 level after the ellipsis.

Student Nurses

High De

Described self as: Absent-minded, complicated, forgetful, headstrong, indifferent, temperamental, withdrawn . . . flirtatious, pleasure-seeking, quarrelsome.

Described by peers as: No adjectives significantly discriminating.

Expected to be described as: Absent-minded, changeable, confused, forgetful, indifferent, irritable, moody, restless, temperamental . . .

Low De

Described self as: Confident . . .

Described by peers as: No adjectives significantly discriminating.

Expected to be described as: Ambitious, confident . . .

Navy Students

High De

Described self as: No adjectives significantly discriminating.

Described by peers as: Confused, impatient, lazy . . . dreamy, immature, interests narrow, irresponsible, self-centered, tactless, careless, foolish, impulsive.

Expected to be described as: Indifferent . . . confused, high-strung, immature, impatient, reckless.

Low De

Described self as: Conscientious, contented, cooperative, gentle, mild, practical, praising, peaceable, pleasant, relaxed, reliable, responsible, steady, tactful, thoughtful . . . clear-thinking, conservative, conventional, capable, cautious, efficient, dependable, moderate, jolly, kind, mannerly, mature, opportunistic, organized, reasonable, resourceful, self-confident, self-controlled, simple, stable, sympathetic, thorough, wholesome.

Described by peers as: Conservative, cooperative, capable, forgiving, generous, helpful, thoughtful, wise . . . alert, clear-thinking, clever, calm, cautious, easy-going, foresighted, frank, initiative, moderate, modest, obliging, kind, mannerly, mature, practical, patient, peaceable, poised, progressive, realistic, reasonable, relaxed, responsible, serious, steady, trusting.

Expected to be described as: Clear-thinking, capable, dependable, patient, peaceable, reliable . . . conservative, cooperative, cautious, honest, planful, reasonable, stable, steady, understanding.

The Navy sample was used in a pilot study and some differences between *S*s and procedures prevent direct comparison of the adjectival content. The Navy sample consisted of 100 *S*s, male and female. They represented approximately the same age range as the student nurses (17–21).

Nevertheless, qualities of indifference and impulsivity appeared in the peer-descriptions of both samples. In the Navy group, some of the adjectives which distinguish the expectations of the high scorers find verification in the peer-descriptions, e.g., "confused," "immature," and "impatient." A correct appraisal of group opinion was made in regard to these characteristics at least. Although the adjective checklist employed in this study has a broad range of description, there may be other self-appraisals on which the high *De* scorers do well.

Summary

The purpose of the study was to test the hypothesis upon which a scale for the detec-

tion of potentially delinquent behavior had been based: that role-taking deficiency and social insensitivity were characteristic of psychopaths. More properly, the investigation concerned the association of social insensitivity with high scores on the scale. The scale had been demonstrated to possess satisfactory screening efficiency, but the operation of the assumed variable, it was suggested, had not been demonstrated.

A total of 204 normal female *Ss* used adjective checklists to describe themselves, described three other acquaintances specified by the investigators, and predicted how they themselves would be described in the composite checklists of the group. A second source of data was obtained by using the peer-nomination technique for designating low and high "insight" *Ss*.

Subjects from the extreme quartiles of the Delinquency (*De*) Scale distribution were compared for predictive accuracy and other variables with the following results and conclusions:

1. Subjects who score high on *De* are significantly less accurate in predicting how they will be described by others than are *Ss* who score low.

2. When peer-nomination ratings of insight were compared with *De* scores, high *De* scorers were rated significantly less insightful than low scorers.

3. High *De* scorers expect to be described by peers in unfavorable terms, and so describe themselves. This expectation is not supported in the descriptions which are actually made by the peers.

4. High *De* scorers tend to use relatively fewer adjectives as they are used by the sample as a whole. Even with conventionality of description held constant, however, they are poorer in predictive accuracy than are other members of the sample.

5. High *De* scorers seem to expect to be misunderstood by their peers.

6. High *De* scorers in a pilot study characteristically expected to be described as confused, immature, and impatient, and were described in these terms by their peers.

If incapacity in role-taking implies a relative inability to understand and predict one's own social stimulus value in a particular setting, the findings of this study support indirectly the theoretical assumption upon which the *De* scale is based. The scale itself apparently discriminates between *Ss* on their ability to "see themselves as others see them."

Received January 15, 1957.

References

1. Gage, N. L., & Cronbach, L. J. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, **62**, 411–422.
2. Gough, H. G. A sociological theory of psychopathy. *Amer. J. Sociol.*, 1948, **53**, 359–366.
3. Gough, H. G. A preliminary guide for the use and interpretation of the California Psychological Inventory. Berkeley: Univer. California Inst. of Pers. Assessment & Res., 1954. (Mimeo.)
4. Gough, H. G. Reference handbook for the Gough Adjective Checklist. Berkeley: Univer. of California Inst. of Pers. Assessment & Res., 1955. (Mimeo.)
5. Gough, H. G. Systematic validation of a test for delinquency. *Amer. Psychologist*, 1954, **9**, 381. (Abstract)
6. Gough, H. G. *Manual for the California Psychological Inventory*. Palo Alto: Consult. Psychologists Press, 1957.
7. Gough, H. G., & Peterson, D. R. The identification and measurement of predispositional factors in crime and delinquency. *J. consult. Psychol.*, 1952, **16**, 207–212.
8. Taft, R. The ability to judge people. *Psychol. Bull.*, 1955, **52**, 1–23.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

Deriving a Measure of Test-Taking Defensiveness

Charles Hanley
Michigan State University

The influence of "defensiveness" and its converse, "plus-getting," on objective tests of personality has been dealt with in a number of ways. One of these is the devising of scales measuring the extent of the behavior and its effect on test scores. Several methods have been used to this end. Hartshorne and May (7), for example, invented items expressing socially undesirable attitudes which presumably would be owned to by most honest people. The *L* scale of the Minnesota Multiphasic Personality Inventory (MMPI) is a current example of this procedure.

Another technique involves the item analysis of responses of normals obtaining normal scores and responses of deviants with normal scores. Items discriminating between these groups are assumed to measure defensiveness. The greater part of the *K* scale of the MMPI was derived in this manner (8).

A procedure introduced by Ruch (9) requires role playing on the part of subjects. Thus Cofer, Chance, and Judson (1) compared the responses of college students who had taken the MMPI under normal conditions with those of students given instructions designed to lead to "faking bad" or "faking good." Item analysis yielded a scale measuring "positive malingering" (defensiveness). In a variation of this procedure, Edwards (3) employed ten judges who were instructed to choose the socially desirable response to each of a pool of MMPI items. Those on which all judges agreed were used as a preliminary scale measuring social desirability effects. Results with this instrument are reported by Fordyce (4).

The present method is in some ways a modification of that proposed by Hartshorne and May. Its rationale first will be described followed by results obtained with an application to the MMPI. It is hoped that the procedure will be useful with any personality test containing a reasonably large number of true-false items.

Rationale of Item Selection

A number of studies have shown a high correlation between the social desirability of attitudes expressed by personality test items and the probability that the items will be endorsed when given in an inventory. Edwards (2), for example, reports a correlation of .87 between endorsement and desirability using a preliminary true-false version of his Personality Preference Schedule. Social desirability of items was rated by judges; probability of endorsement was determined from relative frequency of "true" responses when the inventory was given to subjects under the usual testing conditions. Similar outcomes have been obtained using items from certain MMPI scales (6).

These studies and others reveal that normal subjects tend to endorse socially desirable items, and reject undesirable items when taking personality tests. While it is reasonable to assume that in many cases the endorsements and rejections represent the true state of affairs, an individual lacking these attitudes can obtain a normal score if his answers to items are determined by social desirability rather than personal relevance. Similarly, a normal individual may obtain a deviant score by endorsing socially undesirable items, and rejecting socially desirable items.

To the extent that the relationship between desirability and endorsement can be manipulated by selection of items, it is possible to state conditions under which defensiveness, plus-getting, or both, might be revealed on validating scales.

If with nondefensive subjects the desirability of items is negatively correlated with endorsement, a scale measuring defensiveness can be constructed by keying endorsement of desirable and rejection of undesirable items. When this key is used with ordinary samples, most subjects will obtain low scores, and high scores will be indicative of defensiveness. Items meeting this requirement unfortunately are uncommon.

To measure plus-getting there should be a positive correlation between desirability and probability of endorsement when the latter is based on responses of nonmalingering subjects. Endorsement of undesirable and rejection of desirable items would be keyed in this case, and high scores will reveal malingering when the scale is administered to the usual sample. Typical diagnostic scales, however, generally may meet this prescription; thus high scores can be due either to malingering or genuine deviation. A scale of this kind, therefore, will be useful only if its items are nondiagnostic. It is not likely that an investigator deliberately would include many such items in his original item pool.

A compromise approach is more practical. A scale in which desirability and endorsement are unrelated, when responses are honest, can measure both defensiveness and plus-getting. If endorsement of desirable and rejection of undesirable items are keyed, defensive individuals will obtain high scores and plus-getting subjects low scores. Intermediate scores presumably reflect honest (personally relevant) answers. Since the correlations so far reported between endorsement and desirability although large are not perfect, it is possible, in principle, to locate items, clearly desirable or undesirable, constituting a scale in which this correlation is minimized. Even with Edwards' correlation of .87, inspection of his scattergram (2) reveals that a short defensiveness scale might be derived from his items.

The nondefensive, non-plus-getting group discussed above is hypothetical rather than real. It is likely that in most samples there will be some defensive and a few plus-getting individuals. A scale derived from an honest group will show a positive correlation between desirability and endorsement of items when used with the typical sample. The more defensive the subjects, the higher this correlation will be. If it is assumed, however, that such individuals are in the minority, then the correlation will be small. When, in practice, endorsement values are computed from responses of a mixed group of presumably honest, defensive, and plus-getting subjects, the items may be chosen so as to permit a small, positive correlation between desirability and endorsement.

The assumption that honest subjects are in the majority leads to a deduction regarding the internal consistency of the scale. In the hypothetical honest sample the scale ought to have zero reliability, i.e., the items should not correlate with each other. In a sample of defensive and plus-getting subjects, on the other hand, the internal consistency of the scale should be large; the items ought to have high intercorrelations. In a mixed group with the majority of subjects honest, the internal consistency of the test will be smaller than that usually required for useful measurement. An example may clarify this point. Suppose that the scale has a Kuder-Richardson formula 20 reliability of .80 when given to a mixed sample of 200 defensive and plus-getting subjects. Suppose, further, that when the same scale is administered to 800 honest subjects, the K-R 20 reliability is zero. If these 1,000 subjects are pooled and the reliability recomputed, it will be found to be approximately .39, provided the proportion endorsing each item is the same in both subgroups. In this example, a potentially highly reliable validating instrument will have low reliability when most of the subjects are answering on the basis of personal relevance rather than social desirability, honest answers contributing only error to the validating scale variance. When such a scale used with ordinary samples shows high internal consistency, it might be surmised either that a large proportion of the subjects are not answering honestly, or that the scale reliably measures something in addition to defensiveness and plus-getting.

An Application to the MMPI

Derivation of the Scale

Choice of items will be best when judgments of social desirability are available for

every item in an inventory. In applying the method to the MMPI, the length of the inventory presents a formidable rating problem. For this reason, items presented to judges for desirability ratings were selected in such a manner as to reduce the correlation between endorsement and desirability. One way to minimize the correlation between two variables is to reduce either or both variances. Since variation in desirability is necessary, neutral items not being scorable for defensiveness, restriction of variance in endorsement was the feasible alternative. The items rated by the judges, therefore, were those endorsed by 36 to 64 per cent of Hathaway's normative group of college males and females.[1] There were 53 items of this kind. These, together with ten marker items previously rated by another group of judges, were rated on a nine-point scale of social desirability. Judges were 39 male and 53 female students in introductory psychology classes at Michigan State University. Instructions were those used in an earlier study (6). The social desirability of an item was found by computing its median on the rating scale.

The marker items received ratings quite comparable to those found in the earlier study. Median ratings of four or less were taken as undesirable, four to six as neutral, and six or greater as desirable. Twenty-four of the items fell either into the desirable or undesirable categories. To these were added two items judged undesirable but endorsed by the majority of subjects in the earlier study. The 26-item scale thus formed was keyed for agreement with desirable and rejection of undesirable items. The items keyed "true" are (MMPI Booklet Nos.) 79, 111, 160, 228, 248, 264, 296, 461, and 468. Those keyed "false" are: 15, 30, 71, 109, 124, 135, 142, 148, 170, 324, 383, 406, 408, 409, 416, 439, and 444. Interestingly enough, 12 of these appear on the *K* scale, but the scoring of three of them on the experimental scale is opposite to that used for *K*. Three also are on the *L* scale of the MMPI.

The relationship between desirability and endorsement for the 26-item experimental scale was determined by correlating desirability ratings with endorsement values computed from Hathaway's normative college data. The product-moment coefficient is .27. When endorsement based on responses of 59 male and 41 female introductory psychology students to the booklet form of the MMPI is substituted for the normative group values, the correlation is .31.[2] Thus, the scale appears suitable as a measure of defensiveness and plus-getting.

Reliability

The internal consistency of the experimental scale was computed using K-R 20 and protocols of the same 59 male and 41 female students. The reliability coefficient is .48. A similar coefficient was computed for *K* and found to be .73. The moderate internal consistency of the experimental scale, however, is in line with the assumption that honest responses predominate in typical groups. If the assumption is correct, the higher reliability of *K* is due to the operation of some additional variable or variables not measured by the experimental scale.[3]

Validity

Only indirect evidence of the validity of the experimental scale was obtained. Since the *K* scale of the MMPI has been the subject of much study in this regard, with some evidence pointing to its validity, it was decided to compare the experimental scale and *K* with re-

Table 1

Raw Score Correlations Among Measures of Defensiveness on the MMPI[a]

	Scale				
Scale	*L*	*K*	*Ex*	*SD*	*Cof*
L	—	.40	.57	.35	.42
K		—	.64	.73	.39
Ex			—	.69	.43
SD				—	.46

[a] $N = 100$; an r of .20 is significant at the .05 level.

[1] These percentages are approximate, because no correction was made for normative subjects who used the "?" category.

[2] The author is indebted to Dr. Frederick C. Rockett for making available these MMPI protocols collected in another investigation.

[3] A variable of this kind could be the response set "acquiescence."

Table 2

Raw Score Correlations of the Experimental Scale (*Ex*), and *K* with MMPI Scales[a]

Scale	Scale								
	F	*Hs*	*D*	*Hy*	*Pd*	*Pa*	*Pt*	*Sc*	*Ma*
Ex	−.37	−.44	−.45	.00	−.40	.09	−.65	−.56	−.35
K	−.36	−.34	−.28	.15	−.24	.07	−.69	−.58	−.40

[a] $N = 100$; an r of .20 is significant at the .05 level.

spect to their relationship to the common diagnostic scales of the inventory and to other measures of defensiveness. Table 1 presents the correlations among *L*, *K*, Edwards' recent revision of his scale (*SD*), the Cofer, Chance, and Judson scale (*Cof*), and the experimental scale (*Ex*). All correlations are positive, and, with the exception of the Cofer scale, high. It would be difficult to justify a choice between *K*, *SD*, and *Ex* on these data alone.

If a validating scale is to improve prediction, it is necessary that it correlate with the predicting scales. Correlations between validators and predictors, as the latter now are keyed, may be expected to be negative, i.e., defensiveness will lower scores on predicting scales. The experimental scale and *K* are compared in this respect in Table 2. In Table 3 may be found correlations with the "Obvious" and "Subtle" keys Wiener (11) has developed for several of the MMPI diagnostic scales. (The coefficients in Table 3 generally are higher than the corresponding values in Table 2, a function of the mixed nature of many MMPI diagnostic scales.) Again, it is clear that K and the experimental scale give similar results. In most cases the larger differences in Table 2 may be ascribed to differences in item overlap. The extent of this overlap and the manner in which it is scored are shown in Table 4, similar material on the Obvious and Subtle keys being omitted to save space.

It would be useful if these correlations could be corrected for item overlap. Wheeler, Little, and Lehner (10) have prepared a table based on the independent-elements formula for the Pearson coefficient that indicates the correlation to be expected because of item overlap between various MMPI scales, although they reject its application to the present kind of problem for psychological reasons. Actually, the elements in their situation are test items, and these are not independent when scales have any internal consistency. To correct the present coefficients properly for item overlap would require information regarding the pattern of intercorrelations among all items on the scales, and these data would be tedious indeed to extract. A direct method is necessary to eliminate the effects of item overlap. Results from this procedure will be presented in a later section.

A shortcoming in the foregoing presentation is the absence of material regarding comparative scores of deviant and normal groups. It may be speculated, however, that the less item overlap there is between a diagnostic scale and

Table 3

Raw Score Correlations of the Experimental Scale (*Ex*) and *K* with the Obvious and Subtle Keys of the MMPI[a]

Scale	Obvious Key					Subtle Key				
	D	*Hy*	*Pd*	*Pa*	*Ma*	*D*	*Hy*	*Pd*	*Pa*	*Ma*
Ex	−.61	−.59	−.57	−.43	−.50	.27	.63	.13	.40	.02
K	−.50	−.55	−.60	−.49	−.59	.44	.69	.34	.46	−.01

[a] $N = 100$; an r of .20 is significant at the .05 level.

Table 4

Item Overlap of the Experimental Scale (*Ex*) and *K* with MMPI Scales

Scale	Scoring	MMPI Scales								
		F	*Hs*	*D*	*Hy*	*Pd*	*Pa*	*Pt*	*Sc*	*Ma*
Ex	Same	0	0	1	5	1	2	0	0	2
	Opposite	0	0	4	1	2	2	2	2	1
K	Same	0	0	6	10	7	2	0	1	4
	Opposite	1	0	2	0	1	0	2	0	1

a validating scale, the less likely it will be that the validating scale itself is diagnostic. In the present limited application, it was possible to reduce the amount of overlap compared to that occurring in the *K* scale, as can be seen in Table 4.

The Effect of Acquiescence

Fricke (5) has shown that his measure of acquiescence, the set to respond systematically with "true" or with "false," correlates highly with *K*, and attributed this correlation to the operation of the response set. He further suggests that correlations between *K* and other MMPI scales arise from this set rather than from defensiveness. According to Fricke, correction for acquiescence becomes important when the discrepancy between percentage true and percentage false in a key exceeds 40. This is the case with *K* and Edwards' *SD*, but not with the experimental scale. The correlation of .64 between *K* and the experimental scale could have occurred because of defensiveness and plus-getting, and the correlations of similar magnitude Fricke reports between his measure, *Set T*, and *K* could have arisen because of acquiescence. It is possible, however, that the experimental scale was influenced to some extent by acquiescence, especially when it is noted that the items were chosen from those whose probabilities of endorsement were close to .50. Items of this kind are held to be especially susceptible to the response set.

To settle this problem, recourse was made to empirical evidence. The true-false difference in percentage of keyed responses was set at zero by eliminating eight of the experimental items keyed false (Nos. 15, 135, 142, 148, 324, 383, 409, and 444). Remaining were 18 items, nine true and nine false. These were keyed as before for a measure of defensiveness (*Sx*). The same 18 items were rescored, true responses being keyed. Scored in this manner, the items yielded an "all true" score (*AT*), a measure of acquiescence.

The correlation between *Sx* and *AT* in the 59 males and 41 females is −.11. Thus, as far as the scores represent the effects of individual differences in defensiveness and in acquiescence, the two appear independent. The K-R 20 reliability of *Sx* is .31, while that of *AT* is .32. In spite of the small number of items, both ways of scoring give measures with some internal consistency.

By correlating both scales with MMPI

Table 5

Correlations of the 18-Item Experimental (*Sx*) and *AT* Keys with MMPI Scales—Item Overlap Eliminated[a]

Scale	Key	Scale							
		Hs	*D*	*Hy*	*Pd*	*Pa*	*Pt*	*Sc*	*Ma*
Obvious	*Sx*	—	−.53	−.51	−.50	−.27	—	—	−.38
	AT	—	.05	.10	.20	.20	—	—	.25
Subtle	*Sx*	—	.29	.42	.23	.27	—	—	.07
	AT	—	−.23	−.40	−.08	−.22	—	—	.01
Total	*Sx*	−.38	−.36	−.15	−.34	.00	−.49	−.42	−.28
	AT	.01	.06	−.25	.04	−.04	.26	.20	.22

[a] $N = 100$; an r of .20 is significant at the .05 level.

diagnostic and validating scales, it was possible to arrive at an idea of the effects of social desirability and acquiescence. The correlations with the diagnostic scales are shown in Table 5. These coefficients are free from item overlap; the common items have been deleted from the diagnostic scales. If the two scorings of the experimental scale are accepted as valid, both acquiescence and social desirability affect scores on most of the MMPI scales. This is especially clear when one considers the Obvious and Subtle keys and those diagnostic scales for which no such keys have been developed (because of lack of subtle items).

Ten of the 18 coefficients in Table 5 involving the *AT* scale are statistically significant. Fricke (5) suggests that a true-false discrepancy of 40 per cent or more is a criterion of susceptibility to acquiescence. In ten of the 18 correlations, the true-false discrepancy of the MMPI scale exceeds this value. Nine of the ten significant coefficients occur with these susceptible scales, and in all instances the signs of the coefficients are in the proper direction. Thus, for example, the *Pt* scale has many more true than false items; highly acquiescent subjects would be expected to receive high *Pt* scores, and the empirical correlation of .26 between *Pt* and *AT* supports this prediction.

While acquiescence plays its role, social desirability is at least as important, if not more so, in the present case. All Obvious scales show negative correlations with *Sx*. Its correlations with the Subtle scales are smaller and positive, suggesting that in some cases subtle items are not necessarily free from social desirability.[4]

The *K* scale was explored in similar fashion. The elimination of item overlap between *K* and the 18 items required pruning eight items from *K*, but in spite of this drastic operation, correlations remain with the two keys. The correlation between *K* and *Sx* is .48, while that between *K* and *AT* is − .39, both significant at the .01 level. Edwards' revised scale also was examined in this manner, the single overlapping item being dropped from *SD*. The correlation between *Sx* and *SD* is .55, between *AT* and *SD*, − .24. The better balance of true and false items on *SD*, as compared to *K*, appears to reduce the effects of acquiescence.

[4] An item may be subtle on one scale but obvious on another. This is not true with respect to its social desirability.

Discussion

The results of this study demonstrate the feasibility of constructing a measure of defensiveness and plus-getting, that involves rather less time and expense than usually is required with other methods of item selection. It is not proposed in this paper, however, to offer the experimental scale as a substitute for any MMPI validating variable, nor is it suggested that the validating arsenal of the psychologist has been increased. The MMPI study was undertaken primarily to verify a line of reasoning. The method described appears to work when applied to the MMPI, and it might be expected to be useful when employed with similar kinds of inventories.

A by-product of the study, nevertheless, sheds light on the nature of *K* and other MMPI scales. Both acquiescence and social desirability play their parts, hence some qualification of Fricke's (5) results should be offered. While Fricke selected his items so as to measure acquiescence, it is not now known if these items are neutral in social desirability or whether defensiveness and plus-getting affect scores on his measure.

In the present study, the scoring of the *AT* scale is such that individuals responding defensively obtain average scores, while the *Sx* scoring gives average scores to the highly acquiescent. This is possible because the average item on the experimental scale is endorsed by nearly 50 per cent of the sample. The *AT* and *Sx* scales correlate negligibly; thus measured acquiescence is independent of measured defensiveness. It might be that acquiescence and defensiveness, as measured by ideal instruments, would correlate significantly; there might, for example, be a tendency for acquiescent individuals to be defensive as well. The fact that both scales correlate with other variables in the proper direction, however, indicates that the proposed method is useful. The investigator who develops a measure of defensiveness by applying the present procedure could utilize the same items in a measure of acquiescence. Both keys then could be used to obtain information regarding the sus-

ceptibility of his other scales to these diagnostically extraneous variables. In addition, reference to outside criteria may reveal, as has often been suggested, that the validating scales themselves tap personality dimensions rather than temporary response sets.

Summary

The study was devoted to testing the usefulness of a method for selection of items measuring test-taking defensiveness and plus-getting. It is proposed that a scale in which the social desirability of items is uncorrelated with their probability of endorsement can serve as a validating measure when the items are keyed in terms of the socially desirable response.

A limited application to the MMPI demonstrated that a 26-item experimental scale is similar to the *K* scale with respect to correlations with the main diagnostic scales of the inventory. Significant correlations were found with other measures of defensiveness.

When the experimental scale was corrected for the acquiescence response set, it was found that both acquiescence and defensiveness could be measured by appropriate keying of the same set of items. Correlations of these two keys with other MMPI scales indicated that both defensiveness and acquiescence contribute to the variance on the diagnostic measures.

Received December 18, 1956.

References

1. Cofer, C. N., Chance, June, & Judson, A. J. A study of malingering on the Minnesota Multiphasic Personality Inventory. *J. Psychol.*, 1949, **27**, 491–499.
2. Edwards, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90–93.
3. Edwards, A. L. *Personal Preference Schedule, test manual*. New York: Psychological Corp., 1954.
4. Fordyce, W. E. Social desirability in the MMPI. *J. consult. Psychol.*, 1956, **20**, 171–175.
5. Fricke, B. G. Response set as a suppressor variable in the OAIS and MMPI. *J. consult. Psychol.*, 1956, **20**, 161–169.
6. Hanley, C. Social desirability and responses to items from three MMPI scales: *D*, *Sc*, and *K*. *J. appl. Psychol.*, 1956, **40**, 324–328.
7. Hartshorne, H., & May, M. A. *Studies in deceit*. New York: Macmillan, 1928.
8. Meehl, P. E., & Hathaway, S. R. The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.*, 1946, **30**, 525–564.
9. Ruch, F. L. A technique for detecting attempts to fake performance on a self-inventory type of personality test. In Q. McNemar & Maud A. Merrill (Eds.), *Studies in personality*. New York: McGraw-Hill, 1942.
10. Wheeler, W. M., Little, K. B., & Lehner, G. F. J. The internal structure of the MMPI. *J. consult. Psychol.*, 1951, **15**, 134–141.
11. Wiener, D. N. Subtle and obvious keys for the Minnesota Multiphasic Personality Inventory. *J. consult. Psychol.*, 1948, **12**, 164–170.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

Extraversion, Neuroticism, and Manifest Anxiety[1]

A. W. Bendig
University of Pittsburgh

One of the most recently developed temperament inventories is the Maudsley Personality Inventory (1, 2). This inventory contains 48 trichotomously scored items ("Yes-?-No") with 24 items scored for the two scales included in the MPI: Extraversion and Neuroticism. Eysenck reported (2) that the two scales and split-half reliabilities of .77 and .88 and an intercorrelation of − .05 (N = 400) for his normative group and an intercorrelation of .12 for another smaller group (N = 50). Eysenck has hypothesized (1, p. 50) that Taylor's Manifest Anxiety Scale (4) should show a strong positive correlation with the MPI Neuroticism scale and a smaller negative correlation with the Extraversion scale. The present research was designed to independently assess the internal consistency reliability of the MAS and MPI scales, and also to provide evidence as to the interrelationships among these three scales.

The MAS and MPI were administered to college *Ss* enrolled in a one-semester introductory psychology course. The sample for the reliability study consisted of 145 *Ss* (100 men and 45 women), while the sample used in finding scale intercorrelations included 254 *Ss* (210 men and 44 women). The sex groups were so similar in scale means, variances, reliabilities, and intercorrelations that only results from the combined sex groups are reported here.

Reliability of the MAS scale was estimated by Kuder-Richardson Formula 20, while the variation of the same formula for use with trichotomous items (3) was used with the MPI scales. The reliability coefficients for the Anxiety, Extraversion, and Neuroticism scales were .78, .74, and .84 (N = 145). The intercorrelations among the scales were: Anxiety-Extraversion, − .35; Anxiety-Neuroticism, .77; and Extraversion-Neuroticism, − .20. All three of these coefficients are significant at the .01 level (N = 254).

The reliabilities of the MPI scales are very similar to those reported by Eysenck (2) and are about the same size as the MAS reliability although the MPI scales contain only one-half of the number of items on the MAS. The scale intercorrelations also confirm Eysenck (1) although the intercorrelation between the two MPI scales is slightly higher than those he reports.

Brief Report.
Received May 6, 1957.

[1] An extended report of this study may be obtained without charge from A. W. Bendig, Dept. of Psychology, University of Pittsburgh, Pittsburgh 13, Pa., or for a fee from the American Documentation Institute. Order Document No. 5313, remitting $1.25 for microfilm or $1.25 for photocopies.

References

1. Eysenck, H. J. A dynamic theory of anxiety and hysteria. *J. ment. Sci.*, 1955, **101**, 28–51.
2. Eysenck, H. J. Reminiscence, drive, and personality theory. *J. abnorm. soc. Psychol.*, 1956, **53**, 328–333.
3. Ferguson, G. A. A note on the Kuder-Richardson formula. *Educ. psychol. Measmt*, 1951, **11**, 612–615.
4. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

$F - K$ in a Motivated Group

James Drasgow and W. Leslie Barnette, Jr.
University of Buffalo

"Falsification of response has long been known to be an important factor which limited the validity of the personality inventory or questionnaire" (1, p. 491), and as Gough has observed, "A recurring problem in the use of any personality test is the question of dissembling" (3, p. 408). The Minnesota Multiphasic Personality Inventory (MMPI) contains at least four separate scales (*?*, *L*, *K*, *F*) which directly contribute to evaluating the validity of the scores in the profile. Each one of these four scales assesses validity from a somewhat different approach, but apparently the only promising combination of scales to date is Gough's *F* minus *K* (3).

Several empirical studies have reported on the meaningfulness of the values obtained by subtracting the *K* raw score from the *F* raw score (1, 2, 3, 4, 5). These studies have been especially concerned with establishing optimum cutoff points in the $F - K$ distributions. The distributions that have been studied and reported were obtained from subjects under different conditions, e.g., (*a*) "normal" subjects who were given no particular instructions designed to effect $F - K$, (*b*) normal subjects who were instructed to feign abnormality, (*c*) normals who were instructed to act more normal, and (*d*) patients who were given instructions designed to be analogous to those given to the first three "normal" groups. The most rewarding results have been obtained with condition (*b*). These studies have been well replicated and cutoff scores corroborated. The present study focuses on the unsolved reverse problem of detecting profiles which have been faked to make a "good impression."

In an attempt to discover what precise range of $F - K$ values might be used to discriminate the subjects who were instructed to fake a "good" normal profile, Hunt (4) concluded that more research was needed. Cofer (1) also was unable to find discriminating $F - K$ values; Gough (3), re-working Cofer's data, reiterated Cofer's conclusion.

Why is it that the faked-good profiles have been so difficult to detect? In previous studies with subjects working under fake-good instructions, their motivation is open to question. The choice of motivated subjects by previous researchers appears to have been unfortunate since the probability of finding real differences has been minimized by supplying only instructions to stimulate motivation. The present brief report has therefore focused on this aspect and supplies a group with higher motivation.

Subjects

The University's Vocational Counseling Centre provides a job applicant screening service to business and industry. The job applicants from this service formed the group with which we worked. Some of the applicants were applying for jobs with companies without having been previously associated with the company, while others were old-line company employees competing for promotion to a choice spot. All *Ss* were employed males on their "old" jobs at the time of testing. The jobs for which they were being tested included such titles as foreman, salesman, supervisor, superintendent, and vice-president. The jobs can be seen in a framework of advancement and betterment so that in this society we can reasonably infer an appreciable degree of motivation to get the better job. Many of the men said that they had been working years for the job in question and it was not merely a matter of money.

The total number of "industrial cases" with

MMPIs available for use in the present study was 92. Within this total pool, 66 profiles had scores within the normal range ($T = 30$ to 70), and 26 profiles had one or more scores outside. The normal sample of 66 cases is utilized here, the remaining 26 being held for separate study pending accumulation of more cases.

The mean age of the "normal" job applicant was 34 ± 7 with a range from 20 to 55. Seventy per cent of the group was married; the families averaged two children with a range from zero to four. The mean years of education was 14 ± 2 with a range from 8 to 17. Fifty-seven per cent of the sample had attended college; 29% were college graduates.

The modal person in the sample was a 34-year-old white married male with two children and two years of college. He was currently employed, but trying to get a "better" job.

Results and Discussion

All scores on all MMPI scales on the profile were within the accepted normal range as stated earlier. The mean raw F was 1.6 and the SD was 1.5; the mean raw K was 17.6 with SD 3.2. The difference of -16 for $F - K$ is well beyond the .01 level. Table 1 presents the distribution for this normal sample.

Hunt (4) reported a mean $F - K$ of -11 for the group of navy prisoners who were asked to make a good impression, but he was dissatisfied with this statistic because too many normals also gave this value. One might then expect an $F - K$ of this size as an indication of a "normal" amount of hypocrisy which may be associated with making a good impression in this society. Gough (3) gives -7 to -10 as a modal range within which the majority of normals would fall.

Table 1

$F-K$ Distribution from "Normal" MMPIs

$F-K$	N
−8, −9	2
−10, −11	4
−12, −13	10
−14, −15	13
−16, −17	14
−18, −19	10
−20, −21	10
−22, −23	3

The $F - K$ of -16 found among the job applicants is indirectly corroborated by an earlier study of MacLean *et al.* (5) with nurse candidates in a selection situation. Scores greater than -17 were reported as the probably fake-goods. However, their sample was entirely female and they cautioned against generalizing to dissimilar groups. Judging from the similarity of the mean $F - K$ found in their study and in the present one, the difference between male and female is quite less than obviously exists in other areas.

A further corroborating factor and potential source of explanation for the obtained results appeared in the relationship between a job applicant's $F - K$ and his number of dependents. Because of the restriction in the range of the number of dependents and the non-normal nature of the distribution, a nonparametric correlation technique was used to estimate the association (cf. 6, pp. 202–210). The correlation was .61 and significant beyond the .01 level. Concomitantly, it is of interest to report that the correlation between $F - K$ and age was zero, while that between $F - K$ and education was $-.18$ (Pearsonian r's in both instances).

The relatively high relationship indicated by the .61 could probably be interpreted in a variety of ways. The writers would relate it to the American middle class value of upward social mobility. We postulate that the more dependents the client has, the greater will be his felt personal responsibility and that, as a partial consequence, the more motivated he will be to make a good impression so as to secure the proposed upgrading on the job. It is unfortunate that this statistical test can not be computed from the data of MacLean *et al.* (5).

Summary

Other MMPI studies involving $F - K$ in samples where testees have been requested to fake good are criticized on the ground of inadequate motivation or felt responsibility. Results are presented, utilizing 66 normal MMPI profiles obtained from clients tested for upgrading where evidence was available for high motivation. The mean $F - K$ index for this

group was − 16. Age and years of education had little or no effect; number of dependents, however, was significantly related to this index. It is proposed that the felt responsibility and upward motivation of these clients are the important factors in producing such elevated $F - K$ indices.

Received January 21, 1957.

References

1. Cofer, C. N., Chance, June, & Judson, A. J. A study of malingering on the Minnesota Multiphasic Personality Inventory. *J. Psychol.*, 1949, **27**, 491–499.
2. Gough, H. G. Simulated patterns on the Minnesota Multiphasic Personality Inventory. *J. abnormal soc. Psychol.*, 1947, **42**, 215–225.
3. Gough, H. G. The *F* minus *K* dissimulation index for the Minnesota Multiphasic Personality Inventory. *J. consult. Psychol.*, 1950, **14**, 408–413.
4. Hunt, H. F. The effect of deliberate deception on Minnesota Multiphasic Personality Inventory performance. *J. consult. Psychol.*, 1948, **12**, 396–402.
5. MacLean, A. G., Tait, A. T., & Catterall, C. D. The *F* minus *K* index on the MMPI. *J. appl. Psychol.*, 1953, **37**, 315–316.
6. Siegel, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

The Edwards Personal Preference Schedule and Social Desirability

Robert E. Silverman
New York University

One of the major sources of error in personality testing is the test-taking attitude of the subject. This variable is most intrusive in tests which measure independent personality factors. If the subject is resistant against revealing any psychological weaknesses, this resistance will confound the entire scale. One of the features of the recently developed Edwards Personal Preference Schedule (PPS) (1) is the attempt to control test-taking attitude by employing a forced-choice technique in which persons are asked to choose between items matched for social desirability value. Edwards (1) reported that this procedure was effective in minimizing the effects of social desirability. He correlated the 15 test variables with two measures of social desirability, the *K* scale of the Minnesota Multiphasic Personality Inventory (MMPI) (2), and a specially constructed social desirability scale based on items from the MMPI. In general, the correlations were low, indicating that social desirability was not a major factor in the PPS scores.

Because of the importance of the social desirability variables, it was thought advisable to repeat the Edwards study using the *K* scale and a second independent measure of social desirability. This latter measure was based on the differences between scores on the Taylor Manifest Anxiety Scale (MAS) (4) and a forced-choice version of the MAS designed by Heineman (FC) (3). It was reasoned that a defensive test-taking attitude would manifest itself in terms of the differences between scores on the transparent "Yes-No" inventory and the FC inventory in which precautions had been taken to minimize the effects of social desirability.

Subjects and Procedure

The total sample consisted of 147 male students enrolled in the general psychology course in the University College of Arts and Sciences of New York University. All tests were administered in the weekly discussion sections to groups of 20 to 25 *Ss*. The *Ss* were told that the tests were a part of a research project and that their cooperation was desired. In addition, they were informed that they could discuss their test findings with a member of the psychology staff.

During the first testing session, an inventory consisting of the MAS and the *K* scale was administered to the 147 *Ss*. Two weeks later, 116 *Ss* were given the PPS, and four weeks from the first test, 98 *Ss* who had taken both tests were given the FC scale. The attrition of *Ss* was due to class absences.

Results and Discussion

The correlations between the PPS variables, the measures of social desirability, and the anxiety scales for the present study and for the Edwards study are presented in Table 1. The relationships in the two studies are similar. The correlations between the personality variables and the *K* scale are all low. In the present study, four *K*-scale correlations are significant at the .05 level or better. Two of these, Autonomy and Aggression, were also significant in the Edwards study. In order to compare the two sets of *K*-scale correlations, they were treated as intensive scores and assigned ranks. The highest positive correlation was given the rank one and the highest negative correlation the rank 15. The rank-order relationship between the findings in the two studies was + .65, significant at the .01 level.

Table 1

Correlations Between the PPS Variables, and Measures of Social Desirability and Anxiety

	Present study				Edwards study		
PPS Variable	*K*	Difference scale	MAS	Forced choice	*K*	Social des.	MAS
1. Achievement	.09	−.24*	.08	−.25*	−.01	.09	−.14
2. Deference	.33**	.19	−.09	−.03	.03	.09	−.08
3. Order	.23*	.10	−.22*	−.14	.16	.19	−.18
4. Exhibition	−.10	−.19	.05	−.25*	−.05	−.08	.18
5. Autonomy	−.24*	−.29**	.05	−.38**	−.26*	−.17	−.09
6. Affiliation	−.08	−.03	.20	.00	.08	−.02	.09
7. Intraception	.09	.10	.07	.15	.16	.06	−.06
8. Succorance	−.10	−.12	.28*	.23*	−.25*	−.32**	.22*
9. Dominance	.06	−.05	−.23*	−.29*	.04	−.02	.10
10. Abasement	.01	.29*	.12	.35**	.08	−.14	.18
11. Nurturance	.05	.06	.18	.31**	.08	−.08	.07
12. Change	−.08	−.02	−.11	−.17	.05	.08	−.07
13. Endurance	.18	.32**	−.28*	−.05	.24*	.32**	−.22*
14. Heterosexuality	−.05	−.11	.06	−.01	−.15	.07	.03
15. Aggression	−.35**	−.25*	.12	−.18	−.33**	−.10	.00

Note.—In the present study, the correlation between the *K* Scale and the Difference Scale was .44**. In the Edwards study, the correlation between *K* and Social Desirability was .63**.

* .05 level of confidence.
** .01 level of confidence.

The correlations between the personality variables and the Difference Scale are also low with five correlations being statistically significant. Autonomy and Aggression are again among the significant correlations. The social desirability scale developed by Edwards produced two significant correlations with the personality variables. One of these, Endurance, was also significantly correlated with the Difference Scale in the present study.

The correlations between the 15 personality variables and the two anxiety scales presented in Table 1 indicate that in the present study four personality variables are significantly correlated with the Taylor MAS, and seven personality variables are significantly correlated with the FC anxiety scale. Edwards found low but significant correlations between the MAS and the personality variables, Succorance and Endurance. In the present study, both variables are correlated with the MAS and one variable, Succorance, is also related to the FC anxiety scale.

A comparison of the means and *SD*'s of the PPS variables for the present study and for the Edwards normative sample of males showed five of the 15 variables to be significantly different from each other. Although in each case the differences are small, the number five is greater than one would expect if only chance were operating ($p < .01$). The five variables, Deference, Affiliation, Dominance, Endurance, and Aggression, suggest the presence of small but possibly systematic differences between the two samples. However, these differences do not seem to affect the relationships between the PPS and social desirability measures and between the PPS and the anxiety scales in the two studies.

On the basis of the present data and the findings of Edwards, it would appear that social desirability does not contribute appreciably to the PPS scores. Only three of the 15 PPS variables show somewhat consistent correlations with social desirability indices. Among these three, Autonomy, Endurance, and Aggression, the highest correlation is − .35 between the *K* scale and Aggression. However, the fact that these variables are correlated with social desirability indices may be interpreted as a clue to the construct validity of these variables in that they all seem to relate, albeit slightly, to a defensive attitude in which the *S* describes himself as a conformist, as persevering, and as unwilling to manifest hostility.

The correlations between the PPS variables and the FC anxiety scale represent some relationships between two scales which minimize test-taking attitude. The predictive validity of neither scale has been established, but the correlations do provide us with some information concerning self-report behavior. There seems to be a constellation of PPS variables that correlates reliably with the FC anxiety scale. According to Edwards (1) these PPS variables are correlated with each other. The constellation includes Achievement, Autonomy, Dominance, Abasement, and Nurturance. Achievement, Autonomy, and Dominance are negatively correlated with the FC anxiety scale, while Abasement and Nurturance are positively correlated. It would seem that persons who obtain relatively high scores on the FC anxiety scale describe themselves as having little achievement orientation, as being conformists, as unwilling to be leaders, as feeling guilty, and as desiring to help others. It is noteworthy that these relationships are revealed only when social desirability factors are minimized.

Summary

The test-taking attitude often referred to as social desirability, and the 15 variables of the Edwards Personal Preference Schedule were compared using two measures of social desirability. The findings support those of Edwards in that social desirability plays only a slight role in influencing some of the PPS scores. Two variables do show a consistent relationship with social desirability and this was interpreted as possible indication of the construct validity of these variables. Correlations were reported between the two anxiety scales and the PPS variables. Some implications of these correlations were discussed.

Received January 9, 1957.

References

1. Edwards, A. L. *Manual for the Edwards Personal Preference Schedule.* New York: Psychological Corp., 1954.
2. Hathaway, S. R., & McKinley, J. C. *Minnesota Multiphasic Personality Inventory; manual.* (Rev. ed.) New York: Psychological Corp., 1951.
3. Heineman, C. E. A forced-choice form of the Taylor Anxiety Scale. *J. consult. Psychol.,* 1953, **17**, 447–454.
4. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.,* 1953, **48**, 285–290.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

The Correlates of Manifest Anxiety in Perceptual Reactivity, Rigidity, and Self Concept[1]

Emory L. Cowen, Fred Heilizer, Howard S. Axelrod, and Sheldon Alexander

University of Rochester

In two preceding papers, data have been presented indicating a lack of relationship between scores on the Taylor Manifest Anxiety Scale (29) and performance on two complex learning tasks: (*a*) stylus maze learning (2) and (*b*) paired associate learning (14). These findings are somewhat in contrast to earlier studies in which high anxious subjects (*Ss*) have been found to perform more poorly than low anxious *Ss* on complex learning tasks involving multiple competing response tendencies (13, 17, 23, 25, 30).

The present report deals with the relationship between *A*-scale response and performance on three personality-cognitive measures; perceptual reactivity to "threat expectancy," problem-solving rigidity, and self-regarding attitudes. It evolves from the same basic design as do its two predecessors (2, 14) and accordingly, involves two methodological extensions to earlier studies in the area. Thus, in order to examine the over-the-range discriminatory power of the *A* scale, we have included a middle anxiety group, consisting of MAS scorers between the 43rd and 56th percentiles. And, secondly, we have utilized the MMPI Lie scale as an additional criterion in order to determine whether differences in performance on the dependent measures are related to variations in *L*-scale score.

For the variable of perceptual reactivity, previous investigations (9, 20) report elevated thresholds for correct report of taboo words as compared to neutral ones. While these studies have been criticized (1, 11) in terms of failing to demonstrate *perceptual* defense, the elevation of thresholds of veridical report for taboo words under these operationally defined test conditions remains as verified data. The suggestion has been made (1) that while these studies may demonstrate intraorganismic defensiveness to noxious stimuli, such defensiveness may more parsimoniously be viewed as occurring in processes other than the perceptual one (e.g., inhibition of verbalization). Apart from the question of whether or not such defensiveness is truly *perceptual* in nature, the thesis has been presented that anxiety is the critical variable underlying the frequently observed elevation of thresholds for noxious stimuli (11, 21). If this is so, it might be anticipated that *Ss* with "built-in" high anxiety would have greater elevation of thresholds for anxiety producing stimuli (taboo words) than would low anxiety *Ss*. The present experimental situation offers a test of this proposition.

With respect to the measure of problem-solving rigidity, there is available a consistent body of research which suggests that there is a relationship between experimentally induced anxiety and problem-solving rigidity (6, 7, 24, 27). It might therefore be expected that there would be a comparable relationship between scale anxiety as measured by the Taylor MAS and rigidity in a problem-solving situation. Examination of this relationship constitutes a second aspect of the present research.

The third area of focus is the relationship between *A*-scale anxiety and self-regarding attitudes. Here, the proposition has been

[1] This study was supported in part by a research grant to the senior author, E. L. Cowen, from the National Institute of Mental Mealth, of the National Institutes of Health of the U.S.P.H.S. (MH 437-C). This is the eighth in a series of papers deriving from this work. Portions of the present report were presented in an address to the E.P.A. at the 1954 meetings in New York City.

made that large self-rating discrepancies from differing frames of reference are correlates of a state of internal incongruity or tension within the individual (3, 5, 26). Accordingly, a correspondence between high anxiety and large self-rating discrepancies might be expected. The third aim of the current investigation is that of examining the relationship between these variables.

Subsequent to the collection of our data, two studies have been reported in the literature which bear directly on the present investigation with respect to the variables of problem-solving rigidity and perceptual reactivity to threat expectancy. Maltzman, Fox, and Morriset (19) administered the Taylor *A* scale to 48 *Ss*, who were divided into above and below the median scorers, and were given the Luchins water-jar rigidity task (18). Significantly more rigid solutions were given by HA *Ss* than by LA *Ss*. This finding is somewhat surprising since their *Ss* were divided at the median and thus, presumably, the majority had *A*-scale scores in what has been considered to be the nonsensitive middle range. An additional datum reported by these investigators was that on an extinction problem requiring a new and direct solution, there was a tendency ($p = .16$) for the HA group to obtain fewer correct solutions in the allotted time.

Bitterman and Kniffin (4) have reported a study dealing with manifest anxiety and "perceptual defense." Extreme HA and LA groups were selected on the basis of response to the Taylor *A* scale. Using a tachistoscopic exposure method, all *Ss* were shown four neutral and four taboo words. Taboo words were found to have significantly higher recognition thresholds than neutral words. However, there were no observed differences with respect to the amount of "perceptual defense" (i.e., taboo word thresholds minus neutral word thresholds) between *Ss* with high and low anxiety.

Procedure

Subjects. One hundred and two freshman men and women were selected from a sample of 276 on the basis of their Taylor MAS scores. Three groups were chosen: High Anxiety (HA)—scores of 27 (90th percentile) and above; Middle Anxiety (MA)—scores of 12–14 (43rd–56th percentiles); Low Anxiety (LA)—scores of 6 (20th percentile) and below. The *Ss* in each group were as follows: HA, 28 (14 male and 14 female); MA, 32 (18 male and 14 female); LA, 42 (28 male and 14 female).[2]

Instruments and procedure. Self-regarding attitudes were measured by the Bills-Vance-McLean Index of Adjustment and Values (3), given in group administration. On this instrument, *Ss* are asked to make a series of three ratings along five-point scales, for each of 49 trait-descriptive adjectives. These ratings are: (*a*) the way *S* sees himself with respect to a given trait (self-concept), (*b*) how much he likes being this way (self-acceptance), (*c*) how much, ideally, he would hope to be this way (ideal self). The discrepancy between the self-concept and ideal self ratings summed without respect to sign yields a fourth measure—the maladjustment index. Presumably, the greater the discrepancy between these two ratings the less stable is the self-concept and the more maladjusted the individual. The method for scoring this test is described elsewhere (8).

One to three weeks later, *Ss* were seen individually for the testing of rigidity and perception on the basis of random assignment to one of three *Es*. Problem-solving rigidity was tapped by the Luchins water jar technique (18), a task which involves the establishment of a single mode of solution to a series of arithmetical problems, following which a problem is given for which a more direct and appropriate solution is available. In the present instance, five set problems were used and then an extinction problem soluble only by a new and direct procedure. The rigidity score, as used by Christie (6), was the time of response to the extinction problem, allowing a maximum of ten minutes.

Perceptual reactivity to threat expectancy was studied by a technique described earlier by Cowen and Beier (9). A list of 71 words, 23 presumably threatening, was read initially, with each *S* being advised that he would later

[2] These figures represent the number of *Ss* who completed the anxiety and self-concept scales. Six of these (3 LA, 2 MA, and 1 HA) failed to complete the rigidity and perception experiments.

Table 1

Mean Scores of *A* Scale and *L* Scale Groups for Perception and Self-Concept

Test Variable	Anxiety[a] groups			Lie[a] groups			
				Total group		Low A only	
	High	Mid.	Low	High	Low	High	Low
1. Perception *D* score (trials)	25.6	16.5	19.5	21.2	18.5	22.1	19.4
2. Bills Col. I	169.9	190.2	202.0	197.5	184.4	208.0	193.0
3. Bills Col. II	160.0	175.9	194.5	189.1	175.3	201.9	183.1
4. Bills Col. III	221.1	222.5	223.5	222.3	219.9	226.1	218.5
5. Bills Col. I–III[b]	55.1 (1.7119)	36.3 (1.5403)	25.8 (1.3721)	29.7	39.9	23.3	30.1

[a] *N*s for the several measures are as given in Footnotes 2 and 3.
[b] Since these data were highly skewed, log transformed scores used to compute the *t* ratios reported in Table 4 are given parenthetically.

be asked to decipher some of these words. The *S*s were later shown a series of 16 booklets, each containing 30 carbon copies of a single five-lettered word, electrically typed in capital letters, and arranged in order from the most blurred copy of the clearest. After two practice words, *S*s were given the 16 test booklets (eight threat and eight neutral), in random order, allowing a maximum of three seconds per page within each booklet. The test series of words is presented elsewhere (9). For each *S*, a discrepancy score was computed based on the total number of trials required for correct report of threat words minus the total number of trials for correct report of neutral words.

The present experiment differs from the Bitterman and Kniffin investigation in several procedural respects: (*a*) the use of the "threat-expectancy" situation (i.e., alerting *S*s to the nature of the threatening stimuli beforehand), (*b*) the use of the successive carbon technique in lieu of tachistoscopic presentation, and (*c*) the increased number of stimulus words.

Results

Table 1 summarizes the means for *A*-scale and *L*-scale groups for the perception and self-concept measures.

Rigidity data. The basic score in the analysis of the rigidity data was the time of response to the extinction problem. These scores ranged from one second to the arbitrarily established maximum of ten minutes, and were highly positively skewed. Accordingly, chi square was used as the basis for statistical analysis of these data. Since there was a natural break in the distribution of time scores at approximately the median for the total sample, this was used as the primary basis for dividing the group. Thus 51 *S*s solved the extinction problem in 75″ or less, while the remaining 45 *S*s required 90″ or more to arrive at a correct solution. Chi-square analyses of the rigidity scores for (*a*) the three anxiety groups, (*b*) high and low lie groups for the total sample, and (*c*) high and low lie groups within the low *A* group,[3] were performed. These analyses indicate that there are no significant or near significant relationships between *A*-scale status and rigidity. Chi squares were also computed using finer cutting points for the time scores (e.g., first to fifth quintiles), for *A*-scale and *L*-scale groups, but in no case did any of these analyses yield differences which approached statistical significance.

Perception data. For the total group, the computed correlation between word frequency and recognition threshold was not significant ($r = .08$), suggesting that under the present experimental conditions differential word fre-

[3] Lie scale analyses for the rigidity and subsequent data are based on a comparison of High Lie (scores of six and above) and Low Lie (scores of three and below) groups. For the total sample the *N*s for HL and LL are 25 and 42 respectively, while the comparable figures within the Low Anxiety group are 15 and 14 respectively.

Table 2

Analysis of Variance for Perception *D* Scores

Source	*df*	m.s.	*F*
Anx.	2	592.3	
Exp.	2	3487.0	
Sex	1	1373.7	
A × E	4	970.6	
A × S	2	1233.5	
E × S	2	548.2	
A × E × S	4	1768.3	3.06*
Within	78	577.7	
Total	95		

* Significant at .05 level.

quency was not a factor in determining response threshold. This finding duplicates an earlier one by Cowen and Beier (9), who used similar experimental conditions.

The mean discrepancy score (total threat trials minus total neutral trials) for the group as a whole was found to differ significantly from zero in the direction of more trials required for threat words. The obtained *t* ratio of 6.24 was significant at beyond the .001 level. This datum, too, is in accord with earlier findings by Cowen and Beier (9), as well as those of Bitterman and Kniffin (4).

A three dimensional analysis of variance [4] for perceptual discrepancy scores was computed, and the results of this analysis are summarized in Table 2. The anxiety × sex × examiner interaction proved to be significant when tested over the within error term. In order to use such a significant triple interaction as an error term, at least one of the variables must be randomly distributed. Both the anxiety and sex variables fail to meet this criterion, and since we were unwilling to assume, *post hoc*, a random selection of *Es*, there is no meaningful way in which tests of the higher order effects in the table as a whole may be interpreted [5] (16). Accordingly, separate analyses of variance were computed for each sex, following the rationale detailed in an earlier paper (14). These data are summarized in Table 3. Neither Anxiety nor the Anxiety-Experimenter interaction reached statistical significance for either sex. The Experimenter main effect, however, is statistically significant for men.

Originally, it was intended that a similar analysis of variance be done for the perceptual *D* scores based on a regrouping of *Ss* in terms of their Lie scale scores. The assumption of random distribution of cell entries, however, could not be fulfilled in this case. It was therefore necessary to compute *t* ratios between HL and LL groups for the total sample and within the low MAS group. Significant or near significant differences between groups

Table 3

Analysis of Variance for Perception *D* Scores

Group and Source	*df*	m.s.	*F*
Men			
Anx.	2	260.75	—
Exp.	2	2456.15	4.06*
A × E	4	829.65	1.37
Within	48	604.72	
Total	56		
Women			
Anx.	2	434.45	—
Exp.	2	615.65	1.13
A × E	4	925.90	1.71
Within	30	542.78	
Total	38		

* Significant at .05 level.

[4] Because there were unequal cell entries, the analysis of variance used is an approximation (15) requiring that the following two assumptions be met: (*a*) homogeneity of variance and (*b*) random distribution of cell entries. Homogeneity of variance was tested by means of Bartlett's test and random distribution of cell entries was tested by a chi-square technique. It was possible here to demonstrate that both of the prerequisite assumptions had been fulfilled.

[5] Since the analysis of variance used here is an approximation method, the precision of which is not known (15), *t* ratios were computed between HA, MA, and LA groups in all combinations. None of these *t*'s approached significance at the .05 c.l. The *t* statistic was also computed eliminating *L*-scale scorers of 7 and above, but the results were not altered by this modification.

It may be noted parenthetically that even if the within variance were used as the error term in testing the significance of higher order effects, only the experimenter main effect would be significant. The *F* ratio for the main effect of anxiety would not approach significance.

Table 4

t Ratios Comparing Self-Concept Scores for *A* Scale and *L* Scale Groups

	Anxiety groups compared						Lie groups compared			
	HA vs. MA		HA vs. LA		MA vs. LA		Total group HL vs. LL		Low A only HL vs. LL	
Measure	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*
Col. I	4.74	.001	8.36	.001	3.55	.001	2.85	.01	3.03	.01
Col. II	2.92	.01	6.72	.001	3.67	.001	2.35	.05	2.27	.05
Col. III	0.47	n.s.	0.95	n.s.	0.40	n.s.	0.90	n.s.	2.17	.05
Col. I–III	4.38	.001	7.46	.001	4.08	.001	2.62	.02	1.85	.10

for perceptual *D* scores were not obtained in either of these analyses.

Self-concept data. As the Bills inventory was group administered, the examiner main effect reported in the preceding analysis of variance is no longer present. The primary analysis of these data consists of *t* ratios comparing the several anxiety groups on each of the four Bills measures.[6] Table 4 summarizes the tests of differences between means for *A*-scale and *L*-scale subgroups. With respect to the *A*-scale data, on three of the four self-concept indices there are highly significant differences among anxiety groups, with the HA group always having less well adjusted self-concept scores than either the MA or LA group, and the MA group, in turn, having poorer self-concept scores than the LA group. On the fourth Bills measure (ideal self), there are no significant or near significant differences among anxiety groups.

Regrouping the data in terms of High and Low Lie-scale scores, both for the total sample and within the Low Anxiety group, significant differences between groups are again observed on the Bills measures. For these comparisons, however, the High "Lie" group always scored in the direction of more adequate self-concept.

Discussion

Rigidity. It seems quite clear that the anxiety and rigidity measures used in the present study are not related. The extent to which this finding is in contrast to the data reported earlier by Maltzman *et al.* (19) is less clear. The latter investigators reported significantly more rigid solutions to critical problems for High *A* *Ss*, as well as a tendency (nonsignificant) for these *Ss* to have longer solution times on an extinction problem. Only the extinction problem criterion was used in the present research, the presumption being that by selecting extreme *A*-scale scorers, instead of simply above and below the median scorers as did Maltzman *et al.* (19), the tendency for their above the median scorers to require more time to solve the extinction problem, would be accentuated to the point of statistical significance. This, of course, did not happen in our study.

It is entirely possible that the extinction time criterion is less reliable and/or discriminating than is the alternative criterion of number of indirect solutions to critical problems which discriminated successfully between *A* scale groups in the study by Maltzman *et al.* (19). While there are no data deriving from the present study which could satisfactorily resolve this issue, it may be justifiable to point out that whatever the fallibilities of the extinction time criterion used here, it has at least been demonstrated to be sensitive and reliable enough to be significantly modifiable by experimentally induced stress (6, 7).

In a recent review of this area, Taylor and McNemar (28), noting a comparability in results between the study by Maltzman *et al.* (19) and one by Cowen (7), raise the following provocative question: "Is experimentally induced anxiety the same as manifest anxiety

[6] Although analyses of variance are entirely feasible for these data, they are presented in this less comprehensive form in interests of space, and because the results for the main effect of concern (anxiety) are quite clear-cut.

as measured by the Taylor A scale?" (28, p. 463). There is, by now, substantial evidence in the current literature to suggest that problem-solving rigidity, operationalized *either* in terms of number of indirect solutions to critical problems, or delayed solution of an extinction problem, increases as a function of experimentally induced anxiety (6, 7, 24, 27). The data of the present study suggest that insofar as the second of these two types of rigidity criteria is concerned, there may be substantial differences between experimentally induced anxiety and *A*-scale anxiety.

Perception. Both with respect to the presence of significantly higher thresholds for threat words in the group as a whole, and in terms of the absence of differences in perceptual reactivity to threat and neutral words among anxiety groups, the present findings directly support those of Bitterman and Kniffin (4). These investigators conclude "that anxiety plays no important role" in producing elevated thresholds to taboo words (4, p. 249). The present authors are somewhat reluctant to accept such a conclusion on several grounds:

1. Bitterman and Kniffin take as their "given" the fact that the Taylor *A* scale is a valid measure of anxiety. They write that "the fault might be said to lie with the criterion of anxiety which was used, but there is now considerable evidence to suggest its validity" (4, p. 249). The evidence referred to stems primarily from a series of learning experiments done at the Iowa laboratory (e.g., 13, 17, 23, 25, 30). To the extent that one is willing to accept such a criterion of validity, recent results from the Rochester laboratory (2, 14) and elsewhere suggest that this inferred type of *A*-scale validity is questionable.

2. It is, of course, possible to define *A*-scale score as an operational measure of anxiety, but this in turn raises the question of the relation of this kind of anxiety to other kinds of anxiety. A recent study by Moffitt and Stagner (22) reports different findings for experimentally induced and *A*-scale anxiety in terms of relationships to five perceptual tasks, leading these authors to question the assumption "that manifest and threat induced anxiety were functionally identical" (22, p. 356).

In a somewhat different area, strikingly discrepant findings are reported by Eichler (10), using experimentally induced anxiety, and Westrope (31) using *A*-scale anxiety, both in relationship to Rorschach response. Specifically, within the area of "perceptual defense" data, there is considerable additional evidence suggesting that experimentally induced anxiety leads to elevation of threshold for reactions to noxious perceptual stimuli (11, 12, 21).

Because of the recurrent demonstration of different results in experiments using *A*-scale anxiety as opposed to experimentally induced anxiety, it is difficult to see how the Bitterman and Kniffin data, or our own data for that matter, bear on the relationship of any generalized concept of anxiety to perceptual reactivity. We would be willing only to conclude that *A-scale response* does not relate to perceptual reactivity to threat.

Self-concept. Insofar as the present series of studies is concerned, the clearest relationships between A-scale response and other performance or personality measures are found with respect to the self-concept indices. Here the relationships are marked; apparently, the responses tapped by these two types of measures are quite similar. It should be noted, however, that this relationship does not involve an overt behavioral measure at either end. It is conceivable, therefore, that it reflects nothing more than a common test-taking technique running through two somewhat transparent and structurally similar self-report inventories. More specifically, what perhaps is being reflected in these results is a common desire by some *Ss* to produce a "culturally good" record on both inventories. One clue to the possible accuracy of this line of reasoning derives from the Lie scale analysis of the self-concept data. Here it was found that HL *Ss* produced significantly better self-concept scores than did LL *Ss*, in exactly the same way that LA *Ss* had better self-concept scores than did HA *Ss*. When anxiety was held constant by the comparison of HL and LL *Ss* within the LA group, the HL *Ss* continued to have more "adjusted" self-concept scores.

Summary

The purpose of the present investigation was to examine the relationship between performance on the Taylor *A* scale and measures of problem-solving rigidity, perceptual reactivity to threat-expectancy, and self-concept.

One hundred and two male and female *Ss* were selected as High, Middle, and Low *A* scale scorers. All *Ss* completed the Bills-Vance-McLean Inventory (self concept) in group administration. Subsequently, 96 of these were given the Luchins water jar task of problem-solving rigidity, and a test of perceptual reactivity to threat expectancy, in individual administration.

No relationships between *A*-scale response and performance on either the rigidity or perceptual tasks were observed. Significant differences among anxiety groups were observed on the self-concept indices, such that Low *A* scale scorers characteristically had more adequate self-regarding attitudes. Interpretation of the latter finding is obscured, in that when the data were regrouped, *Ss* with High "Lie" scores were found to have significantly more adequate self-regarding attitudes than did Low "Lie" scorers.

Received January 7, 1957.

References

1. Allport, F. H. *Theories of perception and the concept of structure.* New York: Wiley, 1955.
2. Axelrod, H. S., Cowen, E. L., & Heilizer, F. The correlates of manifest anxiety in stylus maze learning. *J. exp. Psychol.*, 1956, **51**, 131–138.
3. Bills, R. E., Vance, E. L., & McLean, O. S. An index of adjustment and values. *J. consult. Psychol.*, 1951, **15**, 257–261.
4. Bitterman, M. E., & Kniffin, C. W. Manifest anxiety and "perceptual defense." *J. abnorm. soc. Psychol.*, 1953, **48**, 248–252.
5. Brownfain, J. J. Stability of the self-concept as a dimension of personality. *J. abnorm. soc. Psychol.*, 1952, **47**, 597–606.
6. Christie, J. P. The effects of frustration on rigidity in problem solution. Unpublished doctor's dissertation, Univer. of California, 1949.
7. Cowen, E. L. The influence of varying degrees of psychological stress on problem-solving rigidity. *J. abnorm. soc. Psychol.*, 1952, **47**, 512–519.
8. Cowen, E. L. An investigation of the relationship between two measures of self-regarding attitudes. *J. clin. Psychol.*, 1956, **12**, 156–160.
9. Cowen, E. L., & Beier, E. G. Threat-expectancy, word frequencies and perceptual pre-recognition hypotheses. *J. abnorm. soc. Psychol.*, 1954, **49**, 178–182.
10. Eichler, R. M. Experimental stress and alleged Rorschach indices of anxiety. *J. abnorm. soc. Psychol.*, 1951, **46**, 344–355.
11. Eriksen, C. W. The case for perceptual defense. *Psychol. Rev.*, 1954, **61**, 175–182.
12. Eriksen, C. W., & Browne, C. T. An experimental and theoretical analysis of perceptual defense. *J. abnorm. soc. Psychol.*, 1956, **52**, 224–230.
13. Farber, I. E., & Spence, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.
14. Heilizer, F., Axelrod, H. S., & Cowen, E. L. The correlates of manifest anxiety in paired associate learning. *J. Pers.*, 1956, **24**, 463–474.
15. Johnson, P. O. *Statistical methods in research.* New York: Prentice-Hall, 1949.
16. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
17. Lucas, J. D. The interactive effects of anxiety, failure, and intraserial duplication. *Amer. J. Psychol.*, 1952, **65**, 59–66.
18. Luchins, A. S. Mechanization in problem solving: The effect of Einstellung. *Psychol. Monogr.*, 1942, **54**, No. 6 (Whole No. 248).
19. Maltzman, I., Fox, J., & Morrisett, L. Some effects of manifest anxiety on mental set. *J. exp. Psychol.*, 1953, **46**, 50–54.
20. McGinnies, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
21. McGinnies, E., & Sherman, H. Generalization of perceptual defense. *J. abnorm. soc. Psychol.*, 1952, **47**, 81–85.
22. Moffitt, J. W., & Stagner, R. Perceptual rigidity and closure as functions of anxiety. *J. abnorm. soc. Psychol.*, 1956, **52**, 354–357.
23. Montague, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.
24. Pally, S. Cognitive rigidity as a function of threat. *J. Pers.*, 1955, **23**, 346–355.
25. Ramond, C. K. Anxiety and task as determiners of complex performance. *J. exp. Psychol.*, 1953, **46**, 120–124.
26. Rogers, C. R., & Dymond, Rosalind F. *Psychotherapy and personality change.* Chicago: Univer. of Chicago Press, 1954.
27. Rokeach, M. The effect of perception of time on rigidity and concreteness of thinking. *J. exp. Psychol.*, 1950, **40**, 206–216.
28. Taylor, D. W., & McNemar, Olga W. Problem solving and thinking. *Annu. Rev. Psychol.*, 1955, **6**, 455–482.
29. Taylor, Janet A. A personality test for manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
30. Taylor, Janet A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **45**, 61–64.
31. Westrope, Martha R. Relations among Rorschach indices, manifest anxiety and performance under stress. *J. abnorm. soc. Psychol.*, 1953, **48**, 515–524.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

Barron's Ego-Strength Scale: A Replication of an Evaluation of Its Construct Validity[1]

Arthur S. Tamkin and C. James Klett
Veterans Administration Hospital, Northampton, Massachusetts

A previous paper (2) has evaluated the construct validity of Barron's Ego-Strength Scale (*Es*) (1) by determining its correlations with other reputed measures of ego strength and general psychopathology, and by testing its ability to differentiate degree of pathology as determined by psychiatric diagnosis. The measures of psychopathology were the *F* and the Critical Item (CI) scales of the MMPI. While the findings of significant negative correlations between *Es* and *F*, and *Es* and CI were encouraging for the question of validity, the inability of *Es* to discriminate degree of psychopathology based upon psychiatric diagnosis and the failure to demonstrate a significant relationship to Wechsler IQ scores seemed to constitute a serious challenge to its validity when applied to hospitalized psychiatric patients.

Since the number of subjects previously used was small, consisting of 15 psychotics and 15 psychoneurotics and personality disorders, it was decided to replicate the previous study with larger groups of patients from the same institution. One hundred MMPI protocols of the group form were randomly selected from the files of newly admitted or readmitted psychiatric patients. They were scored for *Es*, CI, and *F*, and data on Wechsler full scale IQ, age, education, and established psychiatric diagnosis were obtained for a major part of the sample. In most of the essential respects, previous findings were corroborated at a somewhat higher level of statistical significance. The coefficient of correlation between *Es* and CI was − .66; between *Es* and *F*, − .56; and between CI and *F*, .80. *Es* again failed to separate the two diagnostic groups at the .05 point. Additionally, it was found that *Es* was significantly correlated with IQ (.32) and education (.45) and uncorrelated with age (− .20). The significant coefficient between *Es* and IQ is contrary to the previously reported finding of no significant correlation based upon a sample of smaller *N*, and it brings the findings into line with those of other investigators who have confirmed the relationship between *Es* and intelligence. While the additional corroboration of this fact is suggestive of construct validity for Barron's scale as a measure of ego strength, the confirmation of the scale's inability to separate diagnostic groups presumably of differential levels of ego strength still suggests caution in the application of the *Es* scale to hospitalized psychiatric patients.

Brief Report.
Received May 6, 1957.

[1] An extended report of this study may be obtained without charge from Arthur S. Tamkin, Veterans Administration Hospital, Northampton, Mass., or for a fee from the American Documentation Institute. Order Document No. 5316, remitting $1.25 for microfilm or $1.25 for photocopies.

References

1. Barron, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, 17, 327–333.
2. Tamkin, A. S. An evaluation of the construct validity of Barron's Ego-Strength Scale. *J. clin. Psychol.*, 1957, 13, 156–158.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

Configurational Analysis of MMPI Profiles of Psychiatric Groups[1]

Earl S. Taulbee
VA Mental Hygiene Clinic, Omaha, and University of Nebraska College of Medicine
and Boyd D. Sisson
Medical College of Georgia, Augusta

Although the interest in "sign" versus "configurational" approach in the interpretation of psychological test data has been concerned primarily with the projective techniques, attention has also been given to response "patterns" with some of the more structured tests. Specifically, several studies have dealt with pattern analysis of MMPI data. A sampling of these results clearly reflects disagreement concerning its value as well as effectiveness in differential diagnosis (1, 2, 3, 4, 5).

In 1952, Sullivan and Welsh (10) reported a technique for an objective configurational analysis of the MMPI. The purpose of the present study is to investigate the usefulness of this type of analysis of the MMPI profiles for differentiating, diagnostically and behaviorally, psychoneurotic and schizophrenic groups.

Methods and Subjects

An objective configurational analysis, essentially that developed by Sullivan and Welsh, was applied to the MMPI records of 40 psychoneurotics, hereinafter referred to as the psychoneurotic criterion group (PnCGp) and 40 hospitalized schizophrenics (SchCGp). The neurotic group consisted of half inpatients and half mental hygiene clinic outpatients. Each of the clinical scales of the MMPI was assigned a number from 1 to 9 according to its position on the psychogram; i.e., *Hs* 1, *D* 2 . . . *Ma* 9. The MMPI profile for each patient was coded, with the scale code numbers arranged in order of descending elevation. Rank-order comparisons of all the scales for each record were made by assigning the higher of two scales a plus and the lower one a minus (e.g., if the *Hs* is higher than *D* by one or more *T* scores, the 1 − 2 comparison will be +, while the 2 − 1 comparison will be −). Application of the analysis is illustrated in Table 1. Scales with the same T score were tabulated as equal, and as the number of them was small they were not included in the significance tests. Chi-square tests of differences between the neurotic and schizophrenic groups were then applied to the sums of the pluses and minuses for each of the scale pairs. For example, Table 1 shows that 25 of the 40 neurotics obtained a + score on the *Hs* (1) > *D* (2) comparison and 15 obtained a − score, as compared to 15 and 23, respectively, for the schizophrenics (the *Hs* and *D* scores were equal in two schizophrenic records), yielding an x^2 of 4.1. The groups were differentiated at the .05 level or less by 18 (out of a possible 36) scale pairs.

The mean ages for the PnCGp and SchCGp were 32.6 years (*SD* 8.2) and 30.7 years (*SD* 5.5), respectively; mean education levels were 11.7 years (*SD* 2.9) and 11.9 years (*SD* 2.7). These differences are not significant.

Cross validation of these scale pairs was then carried out by repeating the above procedure using a new group of 40 neurotics (PnVGp), again with an equal number of in- and outpatients, and a new group of 40 hospitalized schizophrenics (SchVGp). In this cross validation, only the 18 scale-pair comparisons which differentiated the original groups were used instead of the possible 36 comparisons. Age and education were com-

[1] From the Veterans Administration Hospital, and Mental Hygiene Clinic, Omaha, Nebraska.

Table 1

Illustration of the Rank-Order Tabulation of the MMPI Profiles (Subject Pn-1, Profile Code, 213478569; Subject Sch-1, Profile Code, 98(56)74312[a])

	Subjects				Totals				PnCGp vs. SchCGp
	PnCGp		SchCGp		PnCGp		SchCGp		
Scale pairs	Pn-1	Pn-2 . . . 40	Sch-1	Sch-2 . . . 40	+	−	+	−	X^2
1–2 (Hs > D)	−		+		25	15	15	23	4.1 (.05)
1–3 (Hs > Hy)	+		−		27	12	16	22	5.7 (.02)
1–4	+		−		28	12	9	30	17.5 (.001)
1–5									
1–6									
1–7									
1–8									
1–9									
2–3									
. . .									
8–9									

[a] The parentheses indicate that the scales have the same *T* score and that the particular scale pair(s) formed by the enclosed numbers are not tabulated. This accounts for the sum of the pluses and minuses not always being 40 in each group for each scale pair. The total scale pair score for the above two profile codes would be 16 for the neurotic and 0 for the schizophrenic.

parable to the criterion groups. Sixteen of the 18 scale pairs again significantly differentiated the groups. (Prior to obtaining test data for the second schizophrenic group, the PnVGp was compared with the SchCGp and the same 16 scale pairs were significant.) These 16 scale pairs which significantly differentiated the two criterion groups from the two validation groups were then applied to a third group of 50 outpatient neurotics ($PnVGp_2$).

All subjects in this study were males and were VA patients with the exception of 30 validation schizophrenics who were patients in a neighboring state hospital.[2] Comparison of the mean profiles for the two schizophrenic groups reveals that they are not significantly different. The schizophrenics were relatively acute cases and more than half of them were classified as the paranoid type. The 90 outpatients were diagnosed independently of their MMPI results; the extent to which the MMPI contributed to the diagnosis of the remaining patients is not known, but it is believed that in the majority of cases the diagnoses were minimally influenced by the MMPI findings.

[2] Grateful acknowledgment is made to Dr. J. L. Yager, Chief Psychologist at the Omaha VA Hospital, for furnishing the MMPI profiles of the VA hospitalized patients, and to Dr. W. G. Klopfer, Chief Psychologist at the Norfolk (Nebr.) State Hospital, for permission to use the records collected by the senior author while he was on the staff at that hospital.

Results

The 16 scale pairs which significantly differentiated the neurotic group from the schizo-

Table 2

Scale Pairs Differentiating the Neurotic and Schizophrenic Groups

	Scale pair	PnCGp vs. SchCGp		PnVGp vs. SchVGp	
		X^2	p	X^2	p
1.	*Hs* (1) > *Hy* (3)	5.7	.02	11.1	.001
2.	*Pd* (4)	17.5	.001	15.5	.001
3.	*Mf* (5)	8.5	.01	8.0	.01
4.	*Pa* (6)	10.2	.01	22.4	.001
5.	*Pt* (7)	4.2	.05	7.2	.01
6.	*Sc* (8)	13.8	.001	26.6	.001
7.	*Ma* (9)	4.9	.05	14.1	.001
8.	*D* (2) > *Pd* (4)	5.1	.05	13.1	.001
9.	*Pa* (6)	4.4	.05	14.8	.001
10.	*Hy* (3) > *Pd* (4)	10.6	.01	18.3	.001
11.	*Mf* (5)	9.5	.01	6.0	.02
12.	*Pa* (6)	8.8	.01	27.6	.001
13.	*Ma* (9)	6.3	.02	6.4	.02
14.	*Pt* (7) > *Mf* (5)	9.9	.01	5.1	.05
15.	*Pa* (6)	10.6	.01	15.6	.001
16.	*Sc* (8)	4.2	.05	14.3	.001

Table 3

Frequency and Percentage of Scale Pairs Occurring in the Criterion and Validation Groups and Differences in the Total Groups

No. of scale pairs	Criterion		Validation		Valid$_2$	Total		X^2	
	Pn	Sch	Pn	Sch	Pn	Pn	Sch	Total	p
13–16	23 (57.5%)	5 (12.5%)	22 (55%)	3 (7.5%)	38 (76%)	83 (63.8%)	8 (10%)	58.48	<.001
7–12	12 (30%)	16 (40%)	16 (40%)	13 (32.5%)	8 (16%)	36 (27.7%)	29 (36.2%)		
0– 6	5 (12.5%)	19 (47.5%)	2 (5%)	24 (60%)	4 (8%)	11 (8.5%)	43 (53.8%)	53.18	<.001

phrenic group (both criterion and validation groups), and the levels of significance, are presented in Table 2. Of the original 18 scale pairs, the two which failed to meet acceptable significance when applied to the validation groups were (*a*) $1 > 2$ and (*b*) $2 > 5$.

Table 3 presents an interval tabulation of the frequency of scale pairs occurring in the different groups. The upper (13–16 scale pairs) and lower (0–6) ranges represent the most effective cutoff points for differentiating the neurotics and schizophrenics, respectively.

After completion of the above procedures and treatment of the data, 100 of the MMPI profiles (40 out- and 20 inpatient neurotics, and 40 schizophrenics) were randomly selected from the total population and submitted to three experienced clinical psychologists and two fourth-year psychology trainees.[3] These judges were instructed to classify each profile as either neurotic or schizophrenic. The only identifying information available to the judges was the individual age, education, and occupation. They were told that every record was either neurotic or schizophrenic, but were not told how many of each were included. In addition, the six signs reported by Meehl (6) in 1946, as later modified for purpose of objective counting by Peterson (8) in his study on the diagnosis of subclinical schizophrenia, were applied to this same sample of 100 MMPI records classified by the judges. These signs are said ". . . to be commonly employed in differentiating MMPI psychotic profiles from neurotic ones (8, p. 198)." The number and percentage of this sample correctly and incorrectly identified by the scale pairs, Meehl's

[3] The authors are deeply indebted to the following psychologists and trainees who served as judges for the study: Drs. Gordon Filmer-Bennett, W. G. Klopfer, Irving Simos, and Messrs. Ralph Anderson and Robert Scott.

Table 4

Number of Correct and Incorrect Identifications of 100 MMPI Profiles by Scale Pairs, Meehl's Signs and Judges

Group	N	Scale pairs		Meehl's signs		J-I		J-II		J-III		J-IVTr		J-VTr	
		+	−	+	−	+	−	+	−	+	−	+	−	+	−
Neurotic	60	42	18	37	23	39	21	41	19	33	27	44	16	30	30
Schizophrenic	40	36	4	26	14	29	11	26	14	27	13	27	13	25	15
Paranoid*	(21)	20	1	12	9	14	7	12	9	15	6	14	7	13	8
Non-paranoid	(19)	16	3	14	5	15	4	14	5	12	7	13	6	12	7
Total	100	78	22	63	37	68	32	67	33	60	40	71	29	55	45

* Note.—Scale pairs identified the paranoid subgroup better than Meehl's signs and all of the judges (at the .05 level or less, using Fisher's *exact test*).

Table 5

Results of Chi-square Test Between the Judges, Scale Pairs and Meehl's Signs

	Meehl's signs	J-VTr	J-IVTr	J-III	J-II	J-I
Scale pairs	5.41*	11.87**	1.29	7.57**	3.03	2.54
J-I	0.55	3.57	0.21	1.39	0.02	
J-II	0.35	3.03	0.37	1.06		
J-III	0.19	0.51	2.68			
J-IVTr	1.44	5.49*				
J-VTr	1.32					

* Significant at less than .05 level.
** Significant at less than .01 level.

signs, and each of the five judges are shown in Table 4.

The results of the chi-square tests of significance among the judges, scale pairs and Meehl's signs are reported in Table 5. Only two of the judges differed significantly in the efficiency of their judgments, although the difference between others approached significance. The scale pairs correctly classified significantly more than two of the judges and Meehl's signs.

Discussion

The findings of this study indicate that the application of a configurational analysis technique for analyzing MMPI data yields an effective means for differentiating groups of neurotic and schizophrenic patients. We believe that the type of analysis presented helps in understanding and evaluating the behavior processes involved in these patient groups, as well as being a very valuable aid in making a differential diagnosis.

Many of the scales which do not differentiate the groups when their mean absolute elevations are considered become very significant when interscale relationships are investigated. For example, the *Mf* scores are almost identical for the two groups, yet the *Mf* scale is significant in three of the 16-scale pairs; similarly for the *Pd* scale. The writers believe that this can be explained by interpreting the data within a psychodynamic frame of reference. The *Sc* and *Pa* scales reflect the greater disturbance in thought processes, more of a tendency to distort and, in general, the more precarious reality contact of the schizophrenic patients. These patients are attempting to alleviate their anxiety by such defenses as projection (*Pa*) and hyperactivity (*Ma*), as contrasted to the neurotics' greater use of vague somatic complaints, repression and obsessive-compulsive behavior (*Hs, Hy,* and *Pt*). The schizophrenics' relatively higher elevation on the *Pd, Pa, Ma, Sc,* and *Mf* scales (compared to *Hs, D, Hy* and *Pt*) is consistent with the generally accepted view that they tend to be more hostile, asocial, suspicious, impulsive, and less bothered by feelings of anxiety, self-doubt, and less able to show guilt or regret than the neurotics. Also, the schizophrenics have stronger feelings of family and social alienation and rejection. The neurotic's conflict usually takes place within himself whereas the schizophrenic, because of his ego's weaker ties with the external world, is more rebellious against conventional practices and expresses many of his emotions, especially hostility, more openly than does the neurotic. The significance of the *Mf* and *Pd* scales in the scale pairs may be the schizophrenics' relatively less adequate identification with the cultural norm of masculinity and greater disregard for conventional behavior in general.

The mean profiles of the in- and outpatient neurotics are similar and differ significantly only on the *Hs* and *Hy* scales, with the outpatient group having greater elevation on both. The mean schizophrenic profile shows a mild negative slope, with the highest score on the *Sc* scale ($T = 69$) and a rise on the *F* scale, as contrasted to the diphasic pattern with a positive slope of the neurotic group. The most striking difference between the results of the two groups is the lower anxiety and relative absence of *Hs–Hy* defenses mani-

fested by the schizophrenics. The large number of paranoid cases probably accounts for no greater right side (*Pa*, *Sc* and *Ma*) elevation than there is but which may be detected in the relatively higher *Pa–Sc* combination.

It is often very difficult to differentiate, on the basis of the MMPI, the acute paranoid schizophrenic from the hospitalized neurotic. This is true because a defensive patient may keep his *Pa* score well within normal limits and also because undue sensitivity and feelings of persecution may be present in many neurotics. The scale pairs may be a valuable aid in making this differentiation, particularly in a group of schizophrenics such as this where no mean scale score was outside the "normal" range. All of the judges misclassified significantly more of the paranoid schizophrenics as neurotics than did the scale pairs. The scale pairs were consistently more successful than the judges in making the neurotic-schizophrenic differentiation, which adds to the list of studies reported by Meehl (7) showing the superiority of the actuarial over clinical prediction. There was a tendency for successful prediction to be related to experience, but the limited number of judges and the fact that the best and poorest performances were by those least experienced, prohibit the drawing of any conclusions concerning the relationship of experience and profile classification.

Of course, the lack of a normal sample and other psychiatric groups limit the application of these findings. However, it is our opinion that in those cases where there is a problem of differentiating between neurotic and schizophrenic conditions, the cautious application of these scale pairs may be quite useful. If the patient's scale pair score is $\geqslant 13$ or $\leqslant 6$, he is likely to be neurotic or schizophrenic, respectively. Any score falling within the 7–12 range should be considered "indeterminate." (Subsequent research by the senior author reveals that latent schizophrenics, neurological cases and "normals" often score between 7 and 12. This would contraindicate using any scores falling in the indeterminate range to classify an individual as being "more like" a neurotic or schizophrenic.)

Summary

A technique of objective configurational analysis was applied to 210 MMPI profiles (two groups of schizophrenics and three groups of neurotics). Sixteen scale pairs were obtained which significantly differentiated two criterion and three validation groups of psychiatric patients, and cutoff ranges are presented which identify them at a very high level of confidence. Application of the analysis is shown to exceed the differentiating efficiency of three experienced clinical psychologists and two advanced psychology trainees. Discussion of the findings with respect to differential diagnosis and behavior processes is presented.

Received January 24, 1957.

References

1. Aaronson, B. S., & Welsh, G. S. The MMPI as a diagnostic differentiator: a reply to Rubin. *J. consult. Psychol.*, 1950, **14**, 324–326.
2. Gough, H. G. Diagnostic patterns on the Minnesota Multiphasic Personality Inventory. *J. clin. Psychol.*, 1946, **2**, 23–36.
3. Guthrie, G. M. Six MMPI diagnostic profile patterns. *J. Psychol.*, 1950, **30**, 317–323.
4. Hunt, H. F., *et al.* Differential diagnostic efficiency of the MMPI. *J. consult. Psychol.*, 1948, **12**, 331–336.
5. Karson, S., & Freud, S. L. Predicting psychiatric diagnoses with the MMPI. *J. clin. Psychol.*, 1956, **12**, 376–379.
6. Meehl, P. E. Profile analysis of the MMPI in differential diagnosis. *J. appl. Psychol.*, 1946, **30**, 517–524.
7. Meehl, P. E. *Clinical versus statistical prediction.* Minneapolis, Minn.: Univer. of Minnesota Press, 1954.
8. Peterson, D. R. The diagnosis of subclinical schizophrenia. *J. consult. Psychol.*, 1954, **18**, 198–200.
9. Rubin, H. The Minnesota Multiphasic Personality Inventory as a diagnostic aid in a veterans hospital. *J. consult. Psychol.*, 1948, **12**, 251–254.
10. Sullivan, P. L., & Welsh, G. S. A technique for objective configural analysis of MMPI profiles. *J. consult. Psychol.*, 1952, **16**, 383–388.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

Scoring Intelligence on the Lowenfeld Mosaic Test[1]

Malcolm H. Robertson
University of Mississippi

The Lowenfeld Mosaic test is a projective test of personality. As with other projective tests, a difference of opinion exists regarding the use of this test to evaluate intelligence. Thus, Diamond and Schmale (1) concluded that there is no relation between performance on the Mosaic test and IQ. However, McCulloch and Girdner (2) reported a correlation of .43 between estimates of mental age and Stanford-Binet mental age.

The hypothesis of this investigation was that meaningful categories could be developed and that scores on these categories show a predictive relationship to intelligence as measured by the Stanford-Binet test.

Ninety elementary school children were divided into three groups of 30 children each on the basis of their IQ levels, above average, average, and below average. The controlled variables were sex, father's occupation, and grade level. The children had been referred to a school clinic for a variety of reasons.

Four psychologists judged the Mosaic productions of each child with respect to general intellectual level and four independent categories. The categories were Movement, Originality, Complexity, and Gestalt. The scoring was completed on a three-step rating scale. Chi-square values and contingency correlations were computed.

For Complexity and Originality, two of the four correlations were significant. Only one correlation was significant for Movement, and none of the correlations were significant for Gestalt. The average correlations for the four judges for Movement, Complexity, Originality, and Gestalt were .22, .33, .38, and .26 respectively. Correlations between estimates of general intellectual level and Stanford-Binet IQ levels were significant for two of the four judges (.42 and .48).

Interjudge agreement on estimates of general intellectual level and the scoring categories was generally significant statistically though the intercorrelations were not high. Intrajudge consistency in scoring the categories and estimating intellectual level was high, although the categories still appear to measure different aspects of performance.

Thus, the Stanford-Binet seems to measure aspects of intelligence which differ largely from what is tapped by the Mosaic test. Any estimate of the general intellectual level from the Mosaic test performance appears to be based to some extent on such characteristics as the complexity, originality, and organization of the Mosaic productions.

Brief Report.
Received May 10, 1957.

[1] An extended report of this study may be obtained without charge from Malcolm H. Robertson, Student Counseling Service, University, Mississippi, or for a fee from the American Documentation Institute. Order Document No. 5314, remitting $1.75 for microfilm or $2.50 for photocopies.

References

1. Diamond, B. L., & Schmale, H. T. The Mosaic Test: I. An evaluation of its clinical application. *Amer. J. Orthopsychiat.*, 1944, 14, 237–250.
2. McCulloch, T. L., & Girdner, L. B. The use of the Lowenfeld Mosaic Test with mental defectives. *Amer. J. ment. Def.*, 1949, 53, 486–496.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

The Social Desirability Stereotype in a Hospital Population

C. James Klett
Veterans Administration Hospital, Northampton, Massachusetts

In earlier studies (2, 6) of the social desirability scale values of the items appearing in the Edwards Personal Preference Schedule (EPPS) (3), it has been demonstrated that judgments about what constitutes socially desirable and undesirable behavior are remarkably stable from group to group and are relatively independent of such variables as age, education, sex, and socioeconomic status. Subcultural differences have been investigated by Fujita (5) and Lovaas.[1] The present paper extends these observations to include a sample of male veterans hospitalized for neuropsychiatric disorders.

It would be intrinsically interesting if such patients deviated systematically from the social desirability stereotype of the nonhospitalized groups. Furthermore, if such were the case, it could be expected to be a determinant in their performance on the EPPS since it has been shown that even slight differences between groups on the matching variable can affect probability of endorsement in forced-choice tests (6).

The 118 subjects used as judges in this study were, for the most part, drawn from the admissions ward and consisted of manifestly disturbed patients, many of whom were actively delusional and hallucinatory. An additional five patients refused the task. By means of the hospital records, 89 of the patients were classified as psychotic and 29 as nonpsychotic.

The subjects were asked to indicate on a nine-point scale their judgment of the desirability of the behavior denoted by each of the

[1] Personal communication.

Table 1

The Relative Standing of College, High School, and Hospitalized Groups in Terms of the Social Desirability of 13 Psychological Needs

Psychological needs	Hospital vs. College	Hospital vs. High School	College vs. High School
1. Achievement	College	Hospital	College
2. Deference	Hospital*	Hospital	High School
3. Order	Hospital**	Hospital*	High School
4. Exhibitionism	College	High School	High School
5. Autonomy	Hospital	Hospital	College
6. Affiliation	College**	High School*	College
7. Intraception	College*	Hospital	College*
8. Succorance	College	High School	College
9. Dominance	Hospital	Hospital	High School
10. Nurturance	Hospital	Hospital	College
11. Change	College	High School*	High School
12. Endurance	Hospital	Hospital	College
13. Aggression	Hospital*	Hospital	High School*

* The mean of this group is significantly higher (more socially desirable) than the mean of the comparison group at the .05 level of confidence.
** Significance at the .01 level.

140 items scaled by Edwards in his development of the EPPS. Cumulative frequency distributions of the judgments were made separately for each item and the median interval values of the psychotic and nonpsychotic groups were compared. Since there was no essential difference between the two groups in terms of the intervals in which the median judgments of the items fell, the separate distributions were combined and the social desirability scale values of the items were obtained by the method of successive intervals (1, 4). These scale values correlated .88 with the scale values of the college group (2), and .87 with those of the high school group (6). Although these correlations are quite high, both are significantly lower than the correlation of .93 obtained between the college and high school scale values.[2]

In order to determine whether differences among these three distributions of scale values were random or were systematically related to item content, the scale values of each of the three distributions were transformed into relative deviates, and comparisons were made for each of 13 psychological needs included in the 140 items.[3] Each need was represented by ten items. The results of the matched pair analyses of variance are summarized in Table 1, which indicates the relative standing of each group in terms of the social desirability of each need.

Discussion

Although there was considerable agreement in what constituted socially desirable and undesirable behavior between the hospitalized group and the nonhospitalized groups, there was significantly less agreement or adherence to the social desirability stereotype than was found between the two nonhospitalized groups. The hypothesis that this disagreement was in any way related to pathology did not receive support from a comparison of psychotic and nonpsychotic patients, but it might be argued that such a dichotomy does not provide sufficient contrast in degree of pathology, besides being subject to erroneous classification. It is conceivable that the judgments of a severe neurotic might be more seriously affected than those of a borderline schizophrenic who is not so acutely ill.

Perhaps a better test is provided by the comparison of the combined hospital group with the two nonhospitalized groups in terms of individual needs. If the disagreement present was due simply to sampling errors, or errors introduced by generalized factors related to pathology such as inattentiveness, confusion, bizarre responses, etc., it could be expected that such errors would be random and not show up as systematic differences as a function of item content (psychological needs). Such was not the case. As can be seen from Table 1, the hospital group as compared with the nonhospitalized groups indicated, at a significant level, that items relating to Deference, Order, and Aggression were more socially desirable and that items relating to Affiliation, Introception, and Change were more undesirable. It would seem that, in spite of the substantial overall agreement in the social desirability stereotype, there are some systematic differences in terms of individual needs.

In the EPPS, items representing different psychological needs were paired on the basis of their social desirability scale values for a college group, the objective being to minimize the effect of social desirability on the probability of endorsement. When administered to groups having different concepts of what is socially desirable, the item pairs are likely to be less adequately matched and, consequently, the probability of endorsement is more likely to be affected by social desirability. This proved to be the case with the high school group even with very high agreement in social desirability scale values (6). If there were sufficient disagreement in social desirability of a systematic nature, this effect could appear as a difference in the means of the EPPS scales for the two groups. If such were the case, the hospital group could be expected to be higher than the college group on those needs judged by its members to be more desirable, and lower on those needs judged to be more undesirable.

Some confirmation is provided by data on

[2] Since scale values were not available for 6 heterosexuality items in the high school and college group and one additional item was not scalable, these correlations are based on an N of 133.

[3] The EPPS is made up of 15 psychological needs, but abasement and heterosexuality are not included in this analysis because all of the scale values were not available. See manual (3) for definitions of the needs.

the EPPS gathered from a group of 40 carefully diagnosed paranoid schizophrenic males in a veterans' hospital.[4] They had significantly higher means on Deference, Order, and Endurance than did the college normative group and significantly lower means on Exhibitionism, Dominance, and Change. The degree of conformity between these data and the present findings is enhanced by the findings on Affiliation, which was lower in the paranoid group although failing of significance, and by Endurance which was judged to be more socially desirable by the hospital group but just failed to reach significance.

It should not be concluded, however, that differences in the means of subscales of the EPPS from group to group are a function of what the group believes to be socially desirable rather than reflecting differences in relative need strength. It is just as obvious that judgments of what is socially desirable are in part a function of the dominant traits of the judging group as it is obvious that probability of endorsement of an item is a function of the social desirability scale value of the item. The two are inextricably confounded. Edwards' solution to the problem seems to be the best one to date with the limitations that: (*a*) it is impossible in practice to match items perfectly, and (*b*) that in spite of striking stability in the social desirability stereotype, there can be systematic differences from group to group that may influence the results obtained on the EPPS.

Summary

The 140 items originally scaled for social desirability during the development of the Edwards Personal Preference Schedule (EPPS) were rescaled using a group of veterans hospitalized for neuropsychiatric disorders. The results were as follows:

1. There was no essential difference between psychotic and nonpsychotic patients in their judgment of the desirability of items.

2. There was a high degree of relationship between the hospitalized group and the college and high school groups in the scale values of the items.

3. In spite of the high agreement, there were systematic differences in the social desirability of subscales representing psychological needs.

4. There was close conformity between those subscales (needs) showing significant differences in social desirability and the subscales showing significant differences in means on the EPPS for a hospitalized group.

Received January 25, 1957.

References

1. Edwards, A. L. The scaling of stimuli by the method of successive intervals. *J. appl. Psychol.*, 1952, **36**, 118–122.
2. Edwards, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90–93.
3. Edwards, A. L. *Edwards Personal Preference Schedule. Manual.* New York: Psychol. Corp., 1954.
4. Edwards, A. L. *Techniques of attitude scale construction.* New York: Appleton-Century-Croft, 1956.
5. Fujita, B. An investigation of the applicability of the Edwards Personal Preference Schedule to a cultural subgroup. Unpublished master's thesis, Univer. of Washington, Seattle, 1955.
6. Klett, C. J. The stability of the social desirability scale values in the Edwards Personal Preference Schedule. *J. consult. Psychol.*, 1957, **21**, 183–185.

[4] Unpublished data gathered by James Milam. Personal communication.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

The Rorschach Performance of Epileptic Children

Merville C. Shaw
Chico State College
and William M. Cruickshank
Syracuse University

A number of studies involving the Rorschach performance of epileptics have been reported. In general, these studies report certain major differences between epileptic and nonepileptic patients. Rorschach himself (6) stated that epileptics show more *S* responses, color naming, *C* responses, *M* responses, and total number of responses. Piotrowski (5) compiled a list of ten signs which he feels are indicative of cortical involvement, although his study did not deal specifically with epileptic subjects. They include: color naming, response total less than 15, average time per response less than one minute, not more than one *M* response, *F* % less than 70, *P* % less than 25, perseveration, impotence, perplexity, and automatic phrases. Presence of any five of these signs in one record was felt to be indicative of the existence of brain damage.

Altable (1) reported the Rorschach performances of thirty epileptics of widely varying ages and with many different types and degrees of epilepsy. He stated that 70–75% of interpretations were based on *F*, that there was an excess of detail percepts, a scarcity of *W*, an excess of white space percepts, and excessive *D* in relation to *W*. A study by Arluck (2) on sixteen idiopathic epileptics also reported certain differences. His subjects were characterized by a *D-Dd* approach, longer reaction time, decreased number of responses and a sum of *CF* and *C* which exceeded *FC*.

Other studies, including those of Guirdham (3) and Martin (4), have reported differences between epileptic and nonepileptic subjects on the Rorschach test, although the differences found by various authors have sometimes been in contradiction to one another.

Method

A number of factors may possibly account for the sometimes contradictory results obtained in studies of this sort: (*a*) failure to use control groups, (*b*) use of epileptic subjects of varying severity and type, and (*c*) inadequate statistical treatment of the data. The current study was designed to overcome, insofar as possible, these shortcomings. Twenty-five institutionalized idiopathic epileptic children classed as grand mal were individually matched with 25 nonepileptic children from other institutional populations on the basis of age, sex, and intelligence. Children were not included in either group if motor defects, uncorrectable sensory defects, or any history which might conceivably indicate brain damage were present.

The mean chronological age of the subjects in the control group was 14 years 3 months, and in the experimental group, 14 years 2 months. The mean IQ of control subjects in the control group was 81.52; of experimental subjects, 81.16. The 1937 revision of the Stanford-Binet was used as the measure of intelligence.

The Rorschach was then administered to all fifty children. The protocols were scored three times by the senior author. The first scoring came almost immediately after the administration of the test, the second after an interval of approximately a month, and the third after the interval of another month. On the final scoring only three changes of any kind were made. The *t* test for matched groups was used to compare the scores of the experimental and control groups on the various Rorschach scoring categories.

Table 1

Significance of Differences Between Epileptic and Control Groups in Rorschach Location, Determinant, and Content Scores

Score	Epileptic Means	Control Means	Difference Between Means	*t*	*P*
Location					
W	3.24	4.10	.86	1.21	n.s.
D	16.84	13.36	3.48	1.51	n.s.
Dd	2.36	1.28	1.08	1.29	n.s.
Determinant					
F	10.12	8.16	1.96	1.41	n.s.
F−	5.04	3.68	1.36	1.86	.10
FC	1.32	1.48	.16	.48	n.s.
C	.32	.16	.16	.88	n.s.
FV	.72	.32	.40	2.38	.05
FY	1.00	1.32	.32	1.09	n.s.
S	1.08	.52	.56	.85	n.s.
Content					
No. different types of content	8.04	6.00	2.04	2.05	.05
No. different types of secondary content	.92	1.20	.28	.97	n.s.
P	3.32	4.08	.76	1.79	.10
A	9.84	10.24	.40	.28	n.s.
Ad	1.48	1.00	.48	1.88	.10
H	1.88	1.96	.08	.07	n.s.
Hd	2.04	1.04	1.00	1.73	.10
Z score	27.58	28.56	.98	.22	n.s.

Results

Analysis of the location of percepts by use of the *t* test indicated that there were no significant differences between the two groups on any of the three variables involved. These results are summarized in Table 1.

Analysis of the differences between the two groups with regard to determinants was carried out in those cases where the frequencies were high enough to make it a valid procedure. In a number of cases it was not possible to do so because of the low frequencies. In four cases, the means of the two groups were identical, or nearly so, and no computations were carried out. This was true in the case of *CF*, *M*, Sum *C*, and *FT*. In the case of the *CF* and *M* determinants, the means were the same for both groups. For the sum of *C*, the means were 1.92 for the experimental group and 2.00 for the control group. For the *FT* determinant, the mean was .80 for the experimental group and .72 for the control group.

In the case of determinants where a plus or minus scoring is possible the signs were disregarded. It was felt that the nature of the determinant, rather than its quality, was the important factor. The procedure was also desirable from the statistical point of view, to increase the number of frequencies of some of the scorings. The determinants for which *t* was computed include *FV*, *FC*, *FY*, *C*, *F*, *F−*, and *S*. In only one case, the *FV* determinant, was there a significant difference between the two groups, at the .05 level. The results are summarized in Table 1.

Statistical tests were applied to the various types of content responses. The *t* test indicated only one difference at the .05 level on any of the variables tested. This was found on the number of different types of content. Differences significant at the .10 level were found on the *P*, *Ad*, and *Hd* variables.

A summary of the differences found between the two groups on the content responses may be found in Table 1.

Table 2

The Average Rorschach of the Epileptic Group Compared to the Average Rorschach of the Control Group

Type of Response	Average Number of Responses	
	Epileptic Group	Control Group
Response Total	22	19
W	3	4
D	17	14
Dd	2	1
F	10	8
F−	5	4
M	1	1
C	0	0
CF	0	0
FC	1	1
Y	0	0
YF	0	0
FY	1	1
V	0	0
VF	0	0
FV	1	0
T	0	0
FT	1	1
Content	8 types	6 types
H	2	2
Hd	2	1
A	10	10
Ad	1	1
Z score	29	29
Total time	9′32″	12′35″
Time per response	1′16″	57″
Time/1R	10.61″	16.74″
Time/1 color R	12.24″	13.86″
Time/1 black R	9.79″	17.76″
Rejections	1	1
Exp/Bal	1/2	1/2
Approach	W:D:Dd	W:D:Dd
Expected Approach	4/16/2	3/14/2
Popular	3	4
Secondary content	1	1
S	1	1
Sum of *C*	2	2

The various time scores kept throughout the Rorschach test were analyzed. There were no significant differences between the two groups on the time of first response, time of first color response, or total time. The number of rejections was not significantly different in the two groups, nor was the response total of each group significantly different from the other.

It is interesting to see what the pattern of the average Rorschach of the epileptic group looks like, compared to that of the control group. The comparison is presented in Table 2. Fractions have been rounded to the nearest whole number. It is perhaps important to note that without the application of proper statistical tests, differences other than those pointed up by the statistical results would be easily inferred.

Summary

The results of the present study fail to confirm most of the alleged Rorschach indicators of epilepsy. The results may possibly be accounted for in several ways. First, the present study seems to be one of the first in which a rigorous system of matching experimental subjects with control subjects has been used. In previous studies the experimental groups have, for the most part, been compared only with the so-called normal Rorschach pattern. Second, the current study made use of statistical tests of significance, rather than subjective comparisons. Finally, this study was confined to one particular diagnostic category of epilepsy of homogeneous severity, while other studies have tended to lump together all types of epileptic patients without regard to etiology. On the basis of the present study, the Rorschach does not appear to be a useful clinic tool for the differential diagnosis of idiopathic epilepsy.

Recevied December 21, 1956.

References

1. Altable, J. P. Rorschach diagnosis in a group of epileptic children. *Nerv. Child,* 1947, **6**, 23–33.
2. Arluck, E. W. A study of some personality differences between epileptics and normals. *Rorschach Res. Exch.,* 1940, **4**, 154–156.
3. Guirdham, A. The Rorschach test in epileptics. *J. ment. Sci.,* 1935, **81**, 870–893.
4. Martin, A. Rorschach pattern of nondeteriorated epileptics. Unpublished doctor's dissertation, Univer. of Texas, 1947.
5. Piotrowski, Z. On the Rorschach method and its application in organic disturbances of the central nervous system. *Rorschach Res. Exch.,* 1936, **1**, 23–39.
6. Rorschach, H. *Psychodiagnostics.* Berne, Switzerland: Hans Huber, 1942.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

The Validity of Some Current Tests For Organicity

James W. Parker
US Army Hospital, Fort Ord, California

Some of the most difficult problems of diagnosis and evaluation are with brain-injured patients. Consequently, a large number of them, either with known lesions or merely those with suspected lesions, are referred to the clinical psychologist for assistance and evaluation. All too frequently, however, the psychological techniques for investigation of organicity fail to offer significant contributions to the understanding of these patients, and the psychologist must report that his findings are equivocal.

In a previous study dealing with the psychological investigation of brain damage, the writer noted that tactual-kinesthetic perception could be used as an effective technique for diagnosing brain damage (7). In the course of that investigation, data were collected from several other psychological techniques frequently employed as adjuncts in the diagnosis of injury to the brain. In order to throw some light on the diagnostic acuity of these techniques, the present report will present the findings resulting from an analysis of these data.

Subjects

The experimental group consisted of 30 adult male hospitalized patients with definitely diagnosed brain injuries of fairly recent origin. Most of them had been injured within six to ten months of the time of this study. They ranged in chronological age from 21 to 48 years, with a mean age of 25 years. Their mental ages, obtained from the Shipley-Hartford Retreat Scale,[1] ranged from 9 years, 6 months, to 18 years, 7 months, with an average mental age of 14 years, 1 month. Patients with gross sensory and motor defects on the dominant side, neuropsychiatric patients, the noticeably mentally deteriorated, and patients with severe visual defects were not included in the investigation.

The control group of 30 adult male patients who revealed no evidence of brain damage was drawn at random from general medical hospital wards. They ranged in chronological age from 19 to 40 years, with an average age of 23 years. Their average mental age was 15 years, 2 months and ranged from 10 years to 18 years, 7 months. The difference of 1 year, 1 month in average mental ages between the two groups was not significant at the .05 level of confidence.

The importance of a control group consisting of hospitalized patients receiving medical care for a physical disorder has too often been overlooked by investigators. Studying only a brain-injured group and a normal group, it has often been assumed that obtained differences were due to the brain injury. However, this is not necessarily the case. There is the possibility that the obtained differences may be due to any organic condition and, therefore, found not only in the case of *S*s with brain damage.

Procedure

The *S*s were referred to the examination by their ward physicians, and apparently accepted it as a part of the routine procedure designed to facilitate understanding of their condition. The following tests were administered: Bender Visual-Motor Gestalt Test,[2] Shipley-Hartford Retreat Scale, Wechsler-

[1] Manson & Grayson (6) have demonstrated the Shipley-Hartford Retreat Scale to be a reasonably valid measure of intelligence.

[2] The Bender-Gestalt test results are reported in a previous study (7) which disclosed that it differentiated the brain-injured from the uninjured *S*s beyond the .01 level of confidence.

Bellevue block-design subtest, and following a ten-minute rest period, the Weigl-Goldstein-Scheerer Color-Form Sorting Test, and the Wechsler Memory Scale. These tests were administered as prescribed in their respective manuals.

Shipley-Hartford Retreat Scale. If Goldstein is correct in his emphasis on the brain-damaged patient's "loss of the abstract attitude" (3), and if the claim is valid that the Shipley-Hartford Retreat Scale measures impairment of the capacity for abstract thinking (8), then the Shipley-Hartford should be effective in diagnosing brain damage. It was administered to the two groups of *Ss*, and their results compared in order to test the ability of this scale to detect brain damage in this population.

Block-design subtest of the Wechsler-Bellevue Scale. It is said that the block designs of the Wechsler-Bellevue differentiate most clearly between brain-injured patients and normal controls (9). For this reason it was included in the present study, and the scores made by the experimental and control groups were compared. Patients with intracranial pathology may be expected to experience particular difficulty on this test, for as Teuber maintains: "Ability in block-design obviously may be limited by even slight traces of visuospatial agnosia, by motor difficulties, and by breakdown of any one among a host of hypothetical 'higher' processes which may have been tentatively called 'visualization,' 'planning,' 'abstraction,' and the like. The point is that we do not know, in advance, what processes are involved in success or failure on such tests" (9, p. 31f).

Weigel-Goldstein-Scheerer Color Form Sorting Test. According to its authors, "The purpose of this test is to determine whether the subject is able to sort a given variety of differently colored figures according to the categories of color or form" (3, p. 110). They advise that *Ss* suspected of impairment of abstract attitude be given the entire series of tests described in their manual; however, they maintain that "indubitably there is some impairment if the subject fails in only one of the tests" (3, p. 30). Furthermore, these authors state that every normal will without difficulty be able to perform successfully on the color form sorting test, and that these *Ss* will be able to shift in their sorting from color to form, and vice versa. A brain-damaged individual, however, is said to be unable to assume the abstract approach and must ". . . therefore discharge the task on an exclusively concrete level of responding . . . [being] unable to *shift volitionally* from his way of grouping to another way" (3, p. 111). Moreover, *Ss* with disturbance of cortical functioning are said to build patterns by arranging the figures in a definite spatial order, while a normal, in arranging the materials, ". . . will casually 'throw' the forms he chooses in different heaps or piles, not being particular about the spacial position of the individual figures within each heap" (3, p. 111).

In order to determine if these claims would apply to the *Ss* of the present investigation, a comparison of the performances of the two groups was made. Frequency of pattern-building was considered, and a simple scoring system was devised so as to have a quantitative measure of success or failure on the task. Those *Ss* who proved able to sort and shift their method of sorting voluntarily received a score of 2; *Ss* who learned to shift from color to form received a score of 1; and those who neither shifted voluntarily nor learned to shift their method of sorting received a score of 0.

Wechsler Memory Scale. It was decided to include the Wechsler Memory Scale in the present experiment in view of the author's statement that this test ". . . should be useful in detecting special memory defects in individuals with specific organic brain injuries, and may prove of concrete value in the examination of some of the soldiers and sailors returning with head injuries" (10, p. 90). In addition to comparing the two groups on the basis of memory quotients derived from this test, their performances on one particular subtest, visual reproduction, were considered separately. Graham and Kendall (4), studying the performance of brain-damaged patients on a memory-for-designs test, observed that these patients made somewhat lower scores than did a control group.

Results

Shipley-Hartford Retreat Scale. The Shipley-Hartford conceptual quotient, which is said to express the degree of impairment in abstract thinking ability (8), failed to discriminate between the two groups in the present experiment. The brain-injured group had a mean conceptual quotient of 88.3, while the controls average 80.3. This difference of eight points proved to be not significant at the .05 level of confidence, which would suggest either that these brain-injured *Ss* were not experiencing a loss in conceptual thinking ability, or that this scale is actually not sensitive to such deficits. In view of the negative findings of previous experiments (1, 6), the latter possibility is not unlikely.

Wechsler-Bellevue Block-design Subtest. This subtest differentiated the brain-injured from the uninjured *Ss* at a highly significant level. The average raw score for the injured was 17.4, while that of the uninjured *Ss* was 25.0. This difference of 7.6 is significant beyond the .01 level of confidence. If the median score of 21 were used as a cutting score, 69 per cent of the brain-injured *Ss* would fall below this, and 37 per cent of the controls would be included in this lower half. Scores below 16 would appear to be particularly suggestive

of brain damage, for while 30 per cent of the experimental group scored below this, only 3.3 per cent of the control group scored this low. Furthermore, these data suggest that very high scores on this test are rarely encountered among brain-injured *Ss*, for while nearly 50 per cent of the controls scored 27 and above, only 3.4 per cent of the experimental group scored this high.

Weigl-Goldstein-Scheerer Color-Form Sorting Test. With this test scored according to the aforementioned method, a chi-square analysis indicated that the experimental and control groups could not be differentiated on this basis. Not only were they alike in their abilities to sort the stimuli and shift their methods of sorting voluntarily, but no significant differences were noted in their tendencies to build patterns with the test materials. The practice of placing the stimuli one on top of another in their sortings was not considered pattern building in the foregoing analysis; however, this type of performance did differentiate the two groups, for 16 of the control group and only 8 of the brain-injured group did this. There was only one *S* in the control group who neither built patterns nor piled the stimuli one on top of the other, while in the experimental group four *Ss* did neither of these things.

These observations seem opposed to Goldstein's finding that while the uninjured *S* casually "throws" (3) the forms into different heaps, the brain-injured patient's spatial arrangement of the figures cannot be arbitrary. This discrepancy between the results of Goldstein's study and those of the present investigation brings to mind the question raised by Hebb, who maintained that "One must object on logical grounds to Goldstein's emphasis on cases which support his interpretations and his disregard for those which are opposed" (5).

Wechsler Memory Scale. The performances of the two groups of *Ss* on this scale were practically identical. The average memory-quotient obtained by the brain-injured *Ss* was 93.9, and that obtained by the controls was 93.7—a difference of only 0.2. The memory-quotients for the brain-injured ranged between 77 and 126, and for the uninjured between 65 and 126. Similar findings resulted from an analysis of the visual-reproduction subtest of this scale. On this subtest, the brain-damaged patients averaged 8.87, and the controls 9.07—again only a 0.2 difference.

The above results would suggest either that these two groups of *Ss* were equal in memory ability, or that memory-quotients derived from this scale are not valid indices of this ability. This latter possibility seems more plausible, since the intracranial lesions—variously located and, in some instances, quite extensive—in the brain-injured *Ss* would very probably have affected memory ability in this group. Furthermore, Eysenck and H. Halstead (2) found that abilities involved in clinical tests of "memory," such as those included in the Wechsler Memory Scale, were the same as those involved in intelligence tests. If this is true, then the two groups in the present study would be expected to score equally well on "memory" tests, since they were found to be equal in mental ages.

Summary

The purpose of the present study was to compare the results of brain-injured *Ss* with the results of uninjured *Ss* on some commonly employed tests for organicity in order to investigate the diagnostic acuity of these techniques. Sixty hospitalized patients were examined, half having brain injuries of fairly recent origin, and half revealing no evidence of neurological involvement. Using the Shipley-Hartford Retreat Scale, the Weigl-Goldstein-Scheerer Color-Form Sorting Test, the Wechsler-Bellevue block-designs, and the Wechsler Memory Scale, only the Wechsler-Bellevue block-design subtest significantly differentiated the two groups. It was made apparent that techniques often used as aids in diagnosing brain damage may be of little or no benefit in this respect.

Received November 19, 1956.

References

1. Aita, J. A., Armitage, S. G., Reitan, R. M., & Rabinowitz, A. The use of certain psychological tests in the evaluation of brain-injury. *J. gen. Psychol.*, 1947, **37**, 25–44.
2. Eysenck, H. J., & Halstead, H. The memory function. I. A factoral study of fifteen clinical tests. *Amer. J. Psychiat.*, 1945, **102**, 174–179.

3. Goldstein, K., & Scheerer, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).
4. Graham, F. K., & Kendall, B. S. Performance of brain-damaged cases on a memory-for-designs test. *J. abnorm. soc. Psychol.*, 1946, **41**, 303–314.
5. Hebb, D. O. Man's frontal lobes. *Arch. Neurol. Psychiat.*, 1945, **54**, 10–24.
6. Manson, M. P., & Grayson, H. M. The Shipley-Hartford Retreat Scale as a measure of intellectual impairment for military prisoners. *J. appl. Psychol.*, 1947, **31**, 67–81.
7. Parker, J. W. Tactual-kinesthetic perception as a technique for diagnosing brain damage. *J. consult. Psychol.*, 1954, **18**, 415–420.
8. *Shipley-Institute of Living Scale for Measuring Intellectual Impairment*. Manual of directions and scoring key. Hartford: The Institute of Living, 1946.
9. Teuber, H. L. Neuropsychology. In R. E. Harris, J. G. Miller, G. A. Muench, L. J. Stone, H. L. Teuber, & J. Zubin, *Recent advances in diagnostic psychological testing*. Springfield, Ill.: Charles C Thomas, 1950.
10. Wechsler, D. A standardized memory scale for clinical use. *J. Psychol.*, 1945, **19**, 87–95.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

The Relation of the Archimedes Spiral Aftereffect and the Trail Making Test to Brain Damage in Children[1]

Anthony Davids, Louis Goldenberg,
Brown University and Emma Pendleton Bradley Home
and Maurice W. Laufer
Emma Pendleton Bradley Home

The present study represents an attempt to explore further the clinical utility of two tests of organicity that have recently been reported in this Journal. Specifically, this experiment is an attempt to extend and to relate the studies reported by Price and Deabler (4) and by Reitan (5).

Price and Deabler reported that "Organic cases with cortical involvement can be differentiated from nonorganic with high degree of certainty by means of the spiral aftereffect technique" (4, p. 302). Results of another experiment, based on the work of these investigators, led Gallese to conclude, "This study agrees with Price and Deabler that the spiral test is of value in differentiating between patients with organic brain damage and those without" (2, p. 257). In an independent paper, Reitan presented results obtained with the Trail Making Test, suggesting that ". . . this short, inexpensive, and easily administered test may be a fairly valid indicator of certain effects of brain damage" (5, p. 394). These studies were concerned with organic damage in adult patients. In the present study, these two tests were brought to bear upon the problem of differentiating between children with cortical damage and children without known cortical impairment.

[1] We wish to express our gratitude to the staff of the Bradley Home for their assistance and cooperation in the execution of this research. Appreciation is also extended to the following members of the East Providence school department who made their facilities available to us and provided the *Ss* for the normal group: Mr. Edward E. Martin, superintendent; Mr. Otho Smith, assistant superintendent; and Mrs. Louise Favorite, psychologist. Sincere appreciation is likewise extended to Mr. Raymond Holden and Miss Nancy DeWolf of the Meeting Street School for their interest and assistance in testing the cerebral palsy children. And we wish to thank Dr. Ralph M. Reitan for providing us with copies of the Trail Making Test.

Method

Apparatus and test material. The spiral aftereffect test employed an electric motor attached to a shaft to produce the necessary rotation. The motor was housed in a small wooden box, painted dull black, with the shaft protruding through an opening in the front of the box. An Archimedes spiral of 920° or 2½ circuits about the center was painted black against a white background and was glued on a disk of white plastic 8 in. in diameter. A mirror image of this spiral was also constructed and glued on a similar plastic disk in order to give the effect of opposite rotation. Thus, two spirals were constructed, Spiral A which gives normal *Ss* a negative aftereffect of expansion, and Spiral B which gives a negative aftereffect of contraction. The disks on which these spirals were mounted were attached, interchangeably, to the shaft protruding from the black box and were turned by the motor at approximately 78 revolutions per minute.

The Trail Making Test used in this experiment was adapted by Reitan from one of the

performance subtests of the Army Individual Test (6) that has also been published as a separate test by Partington and Leiter (3). It is divided into two parts, each consisting of one sheet of 8½ by 11 paper containing sample material on one side and the test material on the other side. Part A consists of 15 circles distributed over the entire page and numbered 1 to 15. Part B consists of 15 circles, eight numbered from 1 to 8, and seven lettered from A to G. The only difference between this children's form of the Trail Making Test and the adult form used by Reitan (5) is that each part of the latter form of the test contained 25 circles.

Subjects. Three groups of *S*s were employed in this study. One group consisted of 15 children from the Meeting Street School in Providence, Rhode Island, which is a special school for children suffering from cerebral palsy. All of these children were diagnosed cerebral palsy, spastic type, and had known cortical damage. The group was composed of seven boys and eight girls, with a mean age of 10.4 years and a mean IQ of 98, as measured by the Wechsler Intelligence Scale for Children (WISC). A second group consisted of 29 children from the Emma Pendleton Bradley Home in East Providence, Rhode Island, which is a residential treatment center for emotionally disturbed children. The majority of these children were free of signs of organic impairment as determined by medical and psychological examinations, and no children with known cortical damage were included in this sample. This group of psychiatric patients was composed of 23 boys and 6 girls, with a mean age of 10.3 years and a mean IQ of 93, as measured by the WISC. A third group consisted of 24 normal children from the East Providence, Rhode Island public school system. These children had no known history of organic impairment. This group was composed of 12 boys and 12 girls, with a mean age of 10.5 years and a mean IQ of 97, as measured by the California Test of Mental Maturity (Short Form).

Procedure. Individual administration of the two tests required about 20 minutes per *S*, and was conducted during the daytime in a room with good illumination. The first part of the testing session was devoted to the spiral aftereffect test. The procedure and instructions were very similar to those described in detail by Price and Deabler (4). Spiral A, giving normal *S*s a negative aftereffect of expansion, was presented first in half of the cases. While the disk was turning, the *S* was asked, "What does the line appear to be doing?" As soon as the disk was stopped, after 30 seconds of rotation, the *S* was asked, "Now what does the line appear to be doing?" and the answer was recorded. The procedure was repeated with Spiral B, giving normal *S*s a negative aftereffect of contraction. This was followed by two repetitions of presentation of A then B, making a total of 6 trials, whereas Price and Deabler's *S*s received only 4 trials. Each correct perception of the negative aftereffect (normal report) was scored one, while failure to perceive the negative aftereffect was scored zero. Doubtful or equivocal reports were scored ½ and, for all *S*s, fractional scores were raised to whole scores in final computations. Thus, the *S*s received total scores ranging from 0 to 6.

The second part of the testing session was devoted to administering the Trail Making Test. On Part A the *S* was required to connect the circles with a pencil line as quickly as possible, beginning with 1 and proceeding in numerical sequence. Then the *S* was administered Part B, on which he was required to connect the circles, alternating between numbers and letters, taking both series in ascending sequence (i.e., 1 to A to 2 to B, etc.). On both parts of the test, whenever an error was made the examiner pointed it out immediately and requested the *S* to make the necessary correction. Since scores for the two parts of the test were the number of seconds required for completion of the task, errors contributed to scores to the extent that additional time was needed for corrections. The raw scores in seconds were used for statistical analyses.

Results

According to the *t* test, there were no significant age or IQ differences among the three groups. A point biserial *r* was computed between IQ and spiral aftereffect scores within each of the groups, yielding coefficients of + .27 (normal group), + .29 (organic group), and + .17 (psychiatric group). None of these

correlations are statistically significant, indicating no relation between spiral scores and IQ.

Spiral aftereffect scores obtained by *Ss* in each of the three groups, together with percentage computations, are shown in Table 1. It is readily apparent that there are pronounced differences among the groups. A chi-square test of association, relating the normal, organic, and psychiatric categories to three scoring categories on the spiral test (failure, partial, or complete success), yielded a value of 26.76 which is significant beyond the .001 level.

The mean spiral scores and the variances for the normal, organic, and psychiatric groups are presented in Table 2. It is evident that with the normal *Ss* there is little variability and the mean score is close to a perfect score. The psychiatric cases are quite variable, but on the average this group did significantly better than did the organic group. This latter group (organic) not only demonstrated considerable variability, but in general their mean score indicated a definite failure to perceive the aftereffect.

Now let us consider results obtained with the Trail Making Test. For each of the three groups, product-moment correlations were computed between IQ scores and scores obtained on Trail A and on Trail B. The only significant coefficient was found within the normal group, where there is a significant correlation between IQ and performance on Trail B. The magnitude of this relation (− .60) suggests that with further refinement this test might be used as a brief and easily administered measure of intelligence in normal children. Although all of the other coefficients are in the direction indicating a negative relation between intelligence and time required to complete the trails, none of these coefficients are significant. Thus, it must be concluded that within the organic and the psychiatric groups there is no significant relation between IQ and trail scores.

Table 1

Number and Percentage of Each Group Achieving Each Spiral Aftereffect Score

Spiral Aftereffect Score	Group					
	Normal		Psychiatric		Organic	
	n	%	*n*	%	*n*	%
6 (Complete)	20	83	18	62	2	13
5	1	4	2	7	0	0
4	1	4	3	10	1	7
3	1	4	0	0	1	7
2	0	0	0	0	1	7
1	0	0	2	7	0	0
0 (None)	1	4	4	14	10	67

Table 2

Comparison of Spiral Aftereffect Scores Obtained by the Three Groups

Group	Mean	Variance
Normal	5.5	1.9
Psychiatric	4.5	5.3
Organic	1.4	5.1

Groups Compared	*F*	*t*
Normal vs. Organic	2.68*	6.70**
Normal vs. Psychiatric	2.79**	2.10*
Psychiatric vs. Organic	1.04	4.21**

* Significant beyond .05 level.
** Significant beyond .01 level.

Mean scores and variances obtained on both parts of this test by each of the three groups and the significance of these differences are shown in Table 3. These findings indicate quite vividly that the normal *Ss* are less variable and perform significantly better on both parts of this test than do the other two groups.[2] On Part A, the psychiatric *Ss* perform significantly better than do the organic *Ss*, but on Part B there is no difference between these two groups of patients.

Finally, results on the two independent tests were related by means of chi-square analyses in order to see if they were associated significantly. Within both the organic group and the normal group it was impossible to form an adequate dichotomy on the basis of spiral

[2] The significance of the differences between these groups was computed using the formulae and procedures described by Walker and Lev (7, p. 157), which take account of the fact that the variances are heterogeneous. These highly significant differences between the means were also obtained employing the method suggested by Cochran and Cox (1) for use when there are significant differences in the variance of responses.

Table 3

Comparison of Trail Making Test Scores (in seconds) Obtained by the Three Groups

Group	Trail A		Trail B	
	Mean	Variance	Mean	Variance
Normal	19	36	47	277
Psychiatric	25	75	83	3546
Organic	32	163	84	2015

Groups Compared	*F*	*t*	*F*	*t*
Normal vs. Organic	4.53**	3.61**	7.27**	2.92**
Normal vs. Psychiatric	2.08*	2.96**	12.80**	3.19**
Psychiatric vs. Organic	2.17*	1.91*	1.76	.34

* Significant beyond .05 level.
** Significant beyond .01 level.

aftereffect scores and therefore impossible to conduct a valid test of association with trail making scores. In the psychiatric group, which permitted a legitimate chi-square analysis, no association was found between spiral scores and scores on Part A of the trails. When the three subsamples were included in one group of 68 *Ss* and chi square was computed, the result was a nonsignificant coefficient of 1.19.

With Trail B, however, there was significant association ($X^2 = 3.70$, $p = .027$) between the psychiatric *Ss*' performance on this test and their performance on the spiral test. Moreover, when the three groups were combined, a coefficient of 9.61 ($p = .01$) was obtained, indicating that for the overall heterogeneous group, consisting of normals, organics, and psychiatric patients, there was a significant degree of association between *Ss*' performance on these two tests of organic impairment.

Discussion

The results of this investigation clearly confirm those reported by Deabler and Price, Gallese, and Reitan. The present study is actually an extension of these studies, and demonstrated the applicability of these tests for use with children. A further contribution of this research is the fact that both tests, which had previously been employed independently, were used in a unified experiment. Thus, all in all, it was found that the findings of previous investigators could be replicated and extended to children with organic impairment, and that there is significant association between the results derived from the two tests. These findings suggest that both of these instruments possess considerable potentiality as valid procedures for diagnosis of organic impairment in children.

In considering the association between the two tests it is interesting to note that within the psychiatric group Trail A scores were not related to spiral scores, while Trail B scores were related to spiral scores. This finding suggests that Trail A may be more susceptible to impulsivity and that merely connecting numbered circles as rapidly as possible does not provide a measure of organic impairment. That is, it is possible that the *Ss* merely rushed through the task which was cognitively very simple and quite concrete, and that there was no relation to their performance on a test such as the spiral test which was probably less susceptible to impulsivity and more a measure of cortical involvement. When confronted with Trail B, however, the task was cognitively complex and the necessity of having to alternate between numbers and letters may well have been abstract enough to provide a measure of organic impairment rather than of impulsivity. At any rate, on the basis of these findings, it would seem that Trail B is probably a better measure of organicity than is the more simple Trail A test.

It is well known that with severely disturbed children contradictory and inconsistent findings are frequently obtained from the

various conventional physiological and psychological tests of organic impairment. Working with children who are undergoing residential treatment, we plan to conduct additional studies in the attempt to gain further understanding of the validity of these assessment methods and of the role of organicity in emotional disturbances of childhood.

Summary

The spiral aftereffect test and the Trail Making Test were administered to a group of normal children, a group of emotionally disturbed children, and a group of children suffering from cortical damage. These two tests, used separately in unrelated studies, had previously been found to differentiate between adults with and without organic brain damage. In the present study both tests were found to reveal significant differences among the groups, as predicted, and both appear to possess considerable promise as valid methods for assessing cortical impairment.

Received January 4, 1957.

References

1. Cochran, W. G., & Cox, G. M. *Experimental designs*. New York: Wiley, 1950.
2. Gallese, A. J. Spiral aftereffect as a test of organic brain damage. *J. clin. Psychol.*, 1956, **12**, 254–258.
3. Partington, J. E., & Leiter, R. G. *Partington Pathways Test*. Washington, D. C.: Psychol. Serv. Center, 1949.
4. Price, A. C., & Deabler, H. L. Diagnosis of organicity by means of spiral aftereffect. *J. consult. Psychol.*, 1955, **19**, 299–302.
5. Reitan, R. M. The relation of the Trail Making Test to organic brain damage. *J. consult. Psychol.*, 1955, **19**, 393–394.
6. U. S. War Department, The Adjutant General's Office. *Army Individual Test; manual*. Washington: Author, 1944.
7. Walker, Helen M., & Lev, J. *Statistical inference*. New York: Holt, 1953.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

The Tendency of the Dörken and Kral Brain Damage Measure to Score False Positives[1]

Harry L. Saslow and William G. Shipman
Staunton Clinic, University of Pittsburgh Medical School

From neurological principles and 190 case records, Dörken and Kral (1) derived a novel scoring system whereby organicity is indicated by the absence of common Rorschach scores. Their "Organic Deficit Rating" (ODR) has not discriminated as well as other Rorschach scoring systems in two validation studies reported to date. This paper reports the ability of this scale to reject nonorganic functionals tested in a psychiatric outpatient clinic.

Rorschach records were obtained from 108 women, aged 20 through 39, seen for diagnostic evaluation at Staunton Clinic. The cases were randomly selected from the files subject to limitations of age, sex, and absence of organicity in the referral note. Intelligence was normally distributed. The sample was about equally divided between neurotics, psychosomatics, and mild personality disorders. Thirty of the records were independently scored by both authors, and high scoring reliability for the ODR is indicated by the Pearson correlation of .96.

Forty-eight of the 108 nonorganics (44%) satisfied the ODR criteria for organicity. The medical records of these cases were checked by board psychiatrists to see if the patients had a previously undetected organic condition. Two organics and three mental defectives were thus found. With this correction, 42 cases (39%) were misclassified by the ODR. This parallels the findings found by others and stands in contrast to the seven per cent inaccuracy found originally by Dörken and Kral.

The scale seems very sensitive to the variable of intelligence since a Pearson correlation of .52 was found between it and Wechsler-Bellevue IQ. Two-thirds of the cases with IQs below 100 obtained an organic score. There were no false positives among the cases with an IQ above 115.

Since anxiety has been shown to constrict Rorschach performance and the ODR tends to label such constricted persons as organic, a check was made on this relationship. The mention of anxiety in the full psychiatric diagnosis, however, did not apply proportionately to more of those classified organic than to those classified nonorganic. False positives did not cluster significantly in any of the medical or psychiatric diagnostic categories applied to the patients.

In summary, the 39 per cent misclassification of functionals by this organicity scale suggests its doubtful utility in an outpatient setting.

Brief Report.
Received June 5, 1957.

[1] An extended report of this study may be obtained without charge from W. G. Shipman, Staunton Clinic, University of Pittsburgh Medical School, Pittsburgh, Pa., or for a fee from the American Documentation Institute. Order Document No. 5315, remitting $1.25 for microfilm or $1.25 for photocopies.

Reference

1. Dörken, H., Jr., & Kral, A. The psychological differentiation of organic brain lesions and their localization by means of the Rorschach test. *Amer. J. Psychiat.*, 1952, **108**, 764–770.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

Base Rate and the Archimedes Spiral Illusion[1]

Donald W. Stilson,[2] Malcolm D. Gynther, and Boris Gertz
South Carolina State Hospital and the University of South Carolina

Recent studies by Price and Deabler (3) and by Page *et al.* (2) provide information concerning the remarkable sensitivity of the spiral aftereffect as an instrument for the detection of organic brain damage. However, the figures which are reported in these studies do *not* provide evidence for the frequency with which the practicing clinical psychologist can expect to differentiate correctly between cases of organic brain damage and functional disorders, using the report of the spiral aftereffect as a diagnostic criterion. The reason for this, as amply illustrated by Meehl and Rosen (1), is that these investigations have not considered the frequency of organic brain damage cases ("base rate" in the terminology of Meehl and Rosen) in the hospital populations from which they have drawn their experimental samples. It is the purpose of this paper to show the effects of estimates of these base rates on the proportion of correct differential diagnoses which can be anticipated, given the sensitivities reported in the investigations mentioned above. In addition, certain general implications of these findings will be discussed.

In the experiments to be discussed, base rates were estimated in accord with the descriptions of the subject samples given by the investigators. Also, the rates were estimated for resident patients, as this appeared to fit best the description of the subjects used in these studies. Since the experiments were intended to evaluate a diagnostic tool, first admissions probably would have been a more appropriate group from which to sample.

[1] The authors wish to thank Dr. M. Kershaw Walsh for a critical reading of the manuscript.

[2] Now at the University of Colorado.

Base Rates and Previous Experimental Findings

Using Price and Deabler's description (3) of their organic population, and referring to the results of the annual census of mental patients conducted by the Public Health Service in 1951 (4), it is possible to estimate that about 16% [3] of the total VA hospital resident population will be included among the categories listed by these investigators. This value, 16%, provides an estimate of the organic base rate for the population from which Price and Deabler obtained their experimental sample.

The base rate implies that if all the patients in this population were diagnosed as "nonorganic" without any test information whatever, then the diagnosis would be incorrect in 16% of the cases.

Price and Deabler found that 98% of their organic group and only 5% of their nonorganic patient group failed to report the spiral aftereffect on four out of four trials. As Page *et al.* did not use any nonpatient subjects, the normal group of Price and Deabler is not included in order that the analyses of the two studies might be comparable. Their patient sample included 75% known organics and 25% nonorganics, and the aftereffect led to a correct diagnosis for 156 out of 160 cases or 2.5% incorrect diagnoses. If their sample had contained only 16% organic patients (as would be more nearly the case in actual use

[3] The authors cannot guarantee that their base rates are the best possible estimates. From published tables (e.g., 4), it is difficult to decide exactly which diagnostic groups should be included under the label "organic." Although Meehl and Rosen (1) assert that determining base rates is simply a matter of computation, decisions must be made with respect to exclusions and inclusions if one is to select the *appropriate* rate.

of the spiral aftereffect in a VA hospital population), the total number of incorrect diagnoses would have been approximately 4.5%. This percentage was obtained by noting that of the 16% organics in that population, 98% would be detected by failure to report the aftereffect on all four trials, whereas, of the 84% nonorganics in the population, only 5% would fail to report the illusion. Of course, use of the aftereffect still represents a substantial reduction in the percentage of errors over what would be obtained using only the base rates for diagnosis, i.e., a $16.0 - 4.5 = 11.5\%$ reduction in incorrect diagnoses.

In the study of Price and Deabler, report or nonreport of the aftereffect on all trials was the most discriminating criterion for separating organic and nonorganic subjects. In view of this fact, it is surprising that Page *et al.* did not give their results using this criterion; rather, they used report of the illusion on at least one trial as the basis for a "nonorganic" diagnosis. These investigators did not present their results in sufficient detail to permit analyzing them using the criterion found most discriminating by Price and Deabler, but both experiments can be analyzed using report of the illusion on at least one trial as the criterion. Using this criterion, 60% of the organic patients in both studies were correctly identified. However, in the Price and Deabler study, use of this criterion led to correct identification of 100% of the nonorganic group, whereas 85% of the Page *et al.* nonorganic sample was so identified.

Another difference between these studies is that Price and Deabler used two spirals (one clockwise and the other counterclockwise) and two trials with each one. Page *et al.* used only one spiral and three trials. It may be that reporting the aftereffect for both clockwise and counterclockwise spirals improves the capacity of the illusion to discriminate organics from nonorganics. A final feature distinguishing the two experimental procedures is the difference in the size of the spirals. Though the effect of this difference may be negligible, it is a possible influence in determining the observed differences.

The appropriate base rate for the Page *et al.* experiment is difficult to determine accurately since 25% of their organic sample were prefrontal lobotomy cases for whom no frequency figures are given in the census (4). The best estimate of the base rate for the Page *et al.* study appears to be .13. This figure differs from the rate estimated for the Price and Deabler study for two reasons: (*a*) Page *et al.* used different categories of organics than did the earlier investigators, and (*b*) the rates for diagnostic classes are different for the two populations sampled (e.g., cerebral arteriosclerotics compose .064 of the state hospital population and only .036 of the VA hospital population). If no test had been employed and all the patients in this population had been diagnosed "nonorganic," this procedure would lead to incorrect diagnoses in 13% of the cases. Using the 40% organic and 85% nonorganic subjects' reporting the aftereffect on at least one trial (2) as a basis for making the diagnosis, the "misses" calculated using the 13% organic base rate would be 18%. In this instance, use of the Archimedes spiral results in an *increase* in errors (i.e., $18.0 - 13.0 = 5.0\%$ more incorrect diagnoses) as compared to predictions from the base rate alone!

General Implications

From the preceding discussion, it is clear that at least two kinds of diagnostic value can be estimated for a given test. The first, which will be called *sensitivity,* is the proportion of cases of a particular diagnostic class (e.g., "organics") which the test indicates are members of this class. The second, *discrimination,* is the proportion of all cases in a particular population which are correctly diagnosed by the test. Sensitivity is unrelated to base rates in a population whereas discrimination depends greatly on the base rates of the populations being considered. Both of the studies cited here give information concerning the sensitivity of the spiral aftereffect for detecting organic and nonorganic patients. Using these sensitivity figures, this paper has pointed out the discrepancy between sensitivity and discrimination of the spiral aftereffect for patient populations similar to those sampled by the two studies.

In a practical situation, discrimination, not sensitivity, is of most concern. The clinical psychologist in a mental hospital must make

diagnoses within a particular population and the adequacy of any diagnostic instrument he uses must be determined with the base rates for that population taken into account.

For the practicing clinician, an effective psychological test for organicity must be able to distinguish between functional disorders and members of the organic group who present diagnostic problems which cannot be resolved better by means of neurological, biochemical, physiological, or other psychological examinations. If organic patients who are clearly diagnosable by other methods (e.g., paretics) are eliminated from a state hospital population, approximately 6.1% of the remainder will be organics and will be among those referred for differential diagnosis. Thus, a diagnosis of "nonorganic" for all of these patients will be incorrect in 6.1% of the cases. If the sensitivities reported by Page *et al.* are used for diagnosis within this population, then 16.5% of the diagnoses would be incorrect, a 10.4% increase in diagnostic errors! With the higher sensitivities reported by Price and Deabler, the proportions of errors still would be 4.8%, which is nearly as many as the 6.1% errors obtained by merely diagnosing all patients "nonorganic." These results demonstrate dramatically the critical role of population base rates in judging the usefulness of a diagnostic test.

Using report of the aftereffect on at least one trial as the criterion for a diagnosis of nonorganicity, the sensitivity of the Archimedes spiral was found to be greater for nonorganics than for organics in both studies. In the Price and Deabler study, it was possible to identify correctly 100% of the nonorganic subjects using this criterion, and in the Page *et al.* experiment, 85% of nonorganics were correctly classified. On the other hand, only 60% of the organic subjects were so identified in both studies. The implication of these findings is that the test apparently can be used with considerable confidence to diagnose *nonorganics*. This emphasizes the need for developing a test with a high sensitivity for organics to be used in conjunction with the Archimedes spiral.

Conclusions

This paper distinguishes between evaluating the sensitivity and the discrimination of a diagnostic instrument. It shows the critical role played by population base rates in determining the discrimination of a diagnostic test, even when the sensitivities for the test are very high. Examination of the base rates may show that basing diagnostic decisions on a greater amount of valid information may increase the number of incorrect diagnoses that are made. Although this paper has focused on the diagnosis of organics with the Archimedes spiral, the implications are of general import for the use of all diagnostic instruments.

Received June 18, 1957.
Early Publication.

References

1. Meehl, P. E., & Rosen, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychol. Bull.*, 1955, **52**, 194–216.
2. Page, H. A., Rakita, G., Kaplan, H. K., & Smith, N. B. Another application of the spiral aftereffect in the determination of brain damage. *J. consult. Psychol.*, 1957, **21**, 89–91.
3. Price, A. C., & Deabler, H. L. Diagnosis of organicity by means of the spiral aftereffect. *J. consult. Psychol.*, 1955, **19**, 299–302.
4. U. S. Department of Health, Education, and Welfare, National Institutes of Health, Biometrics Branch. Patients in mental institutions 1950 and 1951. *Publ. Hlth. Serv. Publ.*, 1954, No. 356.

Journal of Consulting Psychology
Vol. 21, No. 5, 1957

PSYCHOLOGICAL TEST REVIEWS

New Tests

Bruce, Martin M. *Supervisory Practices Test.* Adult. 1 form. Untimed, (20) min. Test booklet (20¢); key (25¢); manual, pp. 7 ($1.00); specimen set ($1.00). New Rochelle, N. Y. (71 Hanson Lane): Author, 1957.

The *Supervisory Practices Test* consists of 50 multiple-choice items which discriminate between supervisors and nonsupervisors in business and industry. Under development since 1948, the test is based on extensive item analyses and cross validations. Retest reliability is reported as .77 and corrected split-half reliability as .82. Correlations with intelligence are low. Validity is described in terms of the separation of groups, correlations with the ratings of foremen, and the correlation of foremen's scores with their employees' morale. The percentile norms show the distributions of 52 high-level executives, 239 supervisors, and 598 nonsupervisory employees. The manual is very satisfactory in the scope of data it supplies and in the conservatism of its recommendations for the use and interpretation of the test.—*L. F. S.*

Cattell, Raymond B. *The I. P. A. T. Anxiety Scale ("Self Analysis Form").* College-adult. 1 form. Untimed, (10) min. Questionnaire booklet ($3.00 per 25); key (40¢); handbook, pp. 21 ($1.00); specimen set ($2.00). Champaign, Ill.: Institute for Personality and Ability Testing, 1957.

In view of the widespread current interest in the concept of anxiety, the publication of a new scale for its measurement is a noteworthy event. The *I. P. A. T. Anxiety Scale* is a product of its author's extensive studies of the factorial structure of personality. A prominent second-order factor of his *Sixteen Personality Factor Questionnaire* has been identified as anxiety (*J. consult. Psychol.*, 1956, 20, 411–418). The present questionnaire consists of 40 items which best represent the five scales most heavily loaded in the anxiety factor—12 items from O (guilt proneness), 10 from Q_4 (id pressure), 8 from Q_3 (−) (lack of integration), 6 from C (−) (ego weakness), and 4 from L (paranoid insecurity). The number of items from each primary approximates its optimum weight in the general factor of anxiety. The *Anxiety Scale*, therefore, rests mainly on the construct validity obtained from the factor analyses.

There are also empirical validities. The score is reported to have high correlation with an objective test factor (U. I. 24) identified as anxiety. It separates normals from anxiety hysterics: about three-fourths of normals fall below, and about three-fourths of anxiety cases above, a raw score of 35. The corrected split-half reliability of the total anxiety score is reported as from .84 to .91, varying with the dispersion of the sample.

In addition to the total score, several special scores are suggested by the manual. The first 20 items are judged more "cryptic" (subtle) and the second 20 more symptomatic or obvious. The ratio of these two scores is suggested as an overt-covert ratio, showing the degree to which the examinee is aware of his anxiety. Unlike the total score, however, this ratio is supported by no evidence from research. The concept is interesting but its application is hazardous. A score may also be obtained for each primary factor represented in the 40 items. Due to brevity, the part reliabilities are only from .26 to .55, and their interpretation is discouraged.

Norms for the total scores are based on 482 men and 313 women, of whom a large but unspecified number are college students. Approximate corrections for sex and age are stated, but are not defended by adequate data. The manual gives 29 references. It seems evident that the *I. P. A. T. Anxiety Scale* has a sounder conceptual base than other current instruments of its type. Many of the functional properties of its scores remain to be established by future research, which will almost surely be forthcoming.—*L. F. S.*

"... a remarkably good condensation of important basic material for students beginning professional training in clinical psychology . . ."

—from a critical comment by Professor Katherine M. Banham, Duke University, on:

INTRODUCTORY CLINICAL PSYCHOLOGY

By SOL L. GARFIELD, *Chief of the Clinical Psychology Division and Associate Professor at The Psychiatric Institute, University of Nebraska Medical School.*

Here is a critical introduction to clinical psychology, covering diagnosis, psychotherapy, and research. It features a study of the development and current role of clinical psychology as a science and profession, with illustrative examples from clinical files.

1957 *469 pages* *$6.00*

TO BE PUBLISHED SPRING 1958—

PERSONAL ADJUSTMENT: An Approach Through the Study of Healthy Personality

By SIDNEY M. JOURARD.

The aim of this book is to provide a clear, explicit description of the healthy personality, as well as a frame of reference for evaluating mental health. The author compares the normal, healthy, and unhealthy expression of basic personality traits; the factors which determine each trait are studied in detail.

DIFFERENTIAL PSYCHOLOGY Third Edition

By ANNE ANASTASI, *Professor of Psychology, Fordham University.*

Incorporating recent findings in biology, anthropology, and ethnic sociology, the third edition of this widely-used text has been extensively reorganized and condensed in order to clarify important concepts. Like the previous editions, the revision treats the major concepts, methodologies, and data pertaining to individual and group differences in behavior. Emphasis is on critical evaluation of procedure and interpretation.

The Macmillan Company

60 FIFTH AVENUE, NEW YORK 11, N.Y.

THE DEVEREUX FOUNDATION

Since 1940 this Foundation, operated on a nonprofit basis, has been the sponsor of Devereux Schools, Communities and Camps.

With extensive facilities in California, Maine and Pennsylvania, the Foundation is in a favorable position to provide medical, psychiatric, psychological, educational and recreational programs for exceptional children on an individual basis.

Devereux Schools were first established in 1912, and they provide a complete scholastic program for children with emotional disorders or impaired intellectual functioning. The importance of the residential nature of the Schools cannot be overemphasized; the thousands of Devereux alumni now leading constructive adult lives are the criterion of the effectiveness of the Schools' work.

Devereux Communities fill a very real need by their "life-experience" and vocational programs for children, adolescents and young adults with impaired intellectual functions. Separate, self-contained campuses allow for homogeneous groupings in terms of age and social maturity.

Devereux Camps are important summer adjuncts to both the Schools and Communities, providing the opportunity for continuity of treatment during summer months. A wide range of play and study is provided, in addition to medical, psychiatric and other formal therapies.

Professional inquiries should be addressed to John M. Barclay, Director of Development, or Charles J. Fowler, Registrar, Devereux Schools. For western states, address Joseph F. Smith, Superintendent, or Keith A. Seaton, Registrar, Devereux Schools in California, Santa Barbara, California.

HELENA T. DEVEREUX
Director

Professional Associate Directors

ROBERT L. BRIDGEN, Ph.D.
MICHAEL B. DUNN, Ph.D.
EDWARD L. FRENCH, Ph.D.
J. CLIFFORD SCOTT, M.D.

DEVEREUX SCHOOLS	DEVEREUX COMMUNITIES	DEVEREUX CAMPS
Pennsylvania	Pennsylvania	Maine
California	California	Pennsylvania
		California